Natural Genetic Variation in Phenotypic, Reproductive and Gene Expression Responses to Whole Genome Duplication in Arabidopsis thaliana.

Chordekar, Sarah

Award date:
2020

Awarding institution:
University of Bath

Natural Genetic Variation in Phenotypic, Reproductive and Gene Expression Responses to Whole Genome Duplication in *Arabidopsis thaliana*.

Sarah Chordekar

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology & Biochemistry

February 2020

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Dr. Paula Kover, for her guidance, encouragement and constructive critique throughout my research. Thank you for your patience and understanding through the many ups and downs of my PhD journey, and for introducing me to the world of bioinformatics, a career that I am enthusiastically in love with!

Gratitude and memories also extend to late Professor Rod Scott, whose suggestions in the initial stages of my research were very valuable. I would also like to thank my co-supervisor, Dr. Leslie Turner, and internal assessors, Dr. Araxi Urrutia and Professor Jason Wolf for their insights that helped in the development of my project over the years. My appreciation also extends to my external examiner, Professor Andrew Leitch, for feedback that helped improve the presentation of my thesis; Dr. Jaume Pellicer and Dr. Ilia Leitch at the Jodrell Laboratory, RBG, Kew for providing us with flow cytometry services for ploidy analysis; Robert Clayton for impeccable glasshouse support; Dr. Atahualpa Castillo for insights in analysis of RNA-Seq data (Chapter 3) and several project students who helped me with optimising methods or collecting data, notably Heather Philips whose dedicated efforts helped gather phenotypic data (Chapter 2). My PhD roller-coaster was made enjoyable and memorable with the friendship of members of the plant lab, specifically Kate Petty, Dr. Mais Sweiss and Dr. Jane Li. I am grateful to the University of Bath for the financial support through the Graduate School Scholarship and the Department of Biology and Biochemistry and Milner Centre for Evolution for their facilities.

My research journey has not only been one of professional development but also that of invaluable personal growth. My acknowledgements would be incomplete without appreciating the constant support of Dr. Momna Hejmadi who was always available with a warm smile and sound advice anytime I needed it. Thank you for looking out for me, motivating me, and helping me maintain perspective, particularly over the past year.

I would also like to thank Anjula Kaushik and Dr. Abhishek Upadhyay for being like family away from home. A huge shout-out to my 2am friends Shreya Banerjee and Sampada Bhandar, who despite being thousands of miles away were always there for me. My interest in research was sparked by inspirational teachers like Pallavi Furia, Narmada Vasani and Prof. Kevin D'Cruz who taught me that knowledge and learning lie beyond textbooks.

Finally, and most importantly, I would like to thank my family, each of whom have contributed to making me the person I am today. I have imbibed the qualities of determination and relentlessness from my mother, Margalit Chordekar; hard-work and composure from my father, Ramakrishnan Parameswaran; creativity and passion from my aunt, Dr. Prakriti Bhaskar; the ability to think out of the box for solutions from my uncle, P.N.Bhaskar; the ability to critique and look at things objectively from my grandfather, Bhailal Shah; the enthusiasm to continuously learn, try new things and perfect them from my grandmother, Anu Shah; diligence, planning, organisation and dedication from my husband, Saurabh Tripathi. Thank you all for your unconditional love, support, encouragement and belief in everything I have ever wanted to do.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY DATA

ABSTRACT

Polyploidy is a key evolutionary process that has helped shape the genomes of several extant eukaryotes, especially flowering plants. Polyploids display significant phenotypic, transcriptomic and reproductive differences relative to their diploids. Oftentimes, this is attributed to large scale genomic reorganisation over evolutionary timescales in ancient polyploids (paleopolyploids), or genetic novelty in polyploids originating from interspecies hybridization (allopolyploids). However, the effects of genome doubling *per se* in polyploids are yet to be clearly understood. Newly formed polyploids that are formed by the somatic doubling of chromosomes (neoautopolyploids) are a good model to study such effects.

Differences between diploids and neoautopolyploids are frequently linked to changes in nucleotype (cell volume, surface area and ratios of components) brought about by the increase in bulk DNA amounts within the nucleus. While true, is unlikely that nucleotype exclusively explains the responses to increased ploidy. The influence that the genetic background of the plant (genotype) has on the determination of ploidy responses has gained little attention.

We used *Arabidopsis thaliana* as a model to address this gap. Neoautopolyploid *A. thaliana* lines of multiple genotypes were generated synthetically using colchicine. The lines were analysed for various phenotypic traits, transcriptomic changes and reproductive differences relative to diploids. We found that in addition to common nucleotype-related responses, the genotype significantly contributed to the responses to increased ploidy. Genotype-related ploidy effects led to a variation in how phenotypic traits responded to increased ploidy; variation in the number and functions of genes that were differentially expressed between genotypes in response to increased ploidy; and variation in the genetic regions of the chromosome that contributed to seed abortion responses in interploidy hybridization. We also observed that colchicine treatment *per se* influenced ploidy responses, highlighting the need to use appropriate controls when comparing synthetic polyploids to diploids.

Our findings demonstrate that genotype is a key component in determining ploidy responses that is capable of impacting all aspects of plant function – molecular, phenotypic and reproductive.

Chapter 1 : **BACKGROUND**

1.1 Polyploidy in nature

Ploidy refers to the number of sets of homologous chromosomes present in the nucleus of an organism. In nature, organisms spanning a wide range of ploidies are observed. Haploids like *Brevipalpus phoenicis*, a parthenogenetic mite, contain one set of homologous chromosomes (1x) (Weeks *et al.*, 2001). Diploids, like most mammals including *Homo sapiens* (humans), contain two sets of homologous chromosomes (2x) and usually receive a set of chromosomes from each of their parents. Polyploids refer to organisms that have more than two sets of chromosomes in their nuclei (Winkler, 1916). Polyploids can vary in the number of chromosome sets they have; from triploids (3x) like cultivated banana to dodecaploids (12x) like *Spartina anglica* (common cordgrass) (Ainouche *et al.*, 2004). Many important cultivated plants are polyploids: Cotton (*Gossypium hirsutum*) is a tetraploid (4X) (Wang *et al.*, 2019), wheat (*Triticum aestivum)* and coffee (*Coffea Arabica*) are hexaploids (6x) (Hart, 1983), and cultivated strawberries are often octoploids (8x) (Edger *et al.*, 2019). While polyploidy may be observed in some prokaryotic bacteria (Mendell *et al.*, 2008; Markov & Kaznacheev, 2016), it is more prevalent across eukaryotes including diatoms (Parks *et al.*, 2018), fungi (Albertin & Marullo, 2012; Todd *et al.*, 2017), fish (Piferrer *et al.*, 2009), amphibians (Mable *et al.*, 2011), mammals (Gallardo *et al.*, 2006), and particularly in angiosperms (flowering plants) (Leitch & Bennett, 1997; Jiao *et al.*, 2011; Datta *et al.*, 2016).

Although organisms are generally categorized into ploidy levels, they can often possess nuclei or tissues of other ploidies. For example, diploids, that spend most of their lifecycle in a diploid state have haploid gamete cell formation via meiosis for reproduction and have transient tetraploid interphase nuclei during cell division by mitosis. Some organisms can also have biphasic life cycles wherein they alternate generations with different ploidy levels and modes of reproduction within their life cycle (Thornber, 2006).

1.2 Frequency of polyploidy in angiosperms

Several attempts have been made to determine the frequency of polyploidy in angiosperms, however, the estimates have changed over time depending on the approach taken to distinguish diploids from polyploids. Historically, the extent of polyploidy was estimated based on the haploid chromosome number 'x' of an organism. Thus, in the 1960's, Grant (1963) proposed that 47% of all angiosperms were of polyploid origin as they had 14 or more chromosomes. This threshold was considered to be too stringent and was subsequently lowered to 11 chromosomes (Goldblatt, 1980) based on which it was estimated that 70-80% of angiosperms were polyploids. As changes in the chromosomes occur over evolutionary time, using absolute chromosome counts as cut-offs did not necessarily provide accurate estimates of polyploid abundance in plants. Several researchers thereafter used novel approaches to estimate polyploid frequency in nature, like stomatal size (Masterson, 1994), or the transitions between odd and even haploid chromosome numbers (Otto & Whitton, 2000), however, as genomes evolve and change over time following whole genome duplication, accurate determination of polyploid frequency using these methods have not been achievable. Thus, alternatively, instead of estimating the number of species that are currently polyploids, researchers attempted to estimate the number of angiosperm species that were descendants from whole genome duplication events (Soltis *et al.*, 2004). On these lines, using a molecular dating and phylogenetic approach, it is now suggested that all existing seed plants have originated from a polyploid ancestor (Jiao *et al.*, 2011) (Figure 1.1).

The common occurrence of genome duplication events combined with the absence of angiosperm lineages that have not experienced a duplication event can be interpreted as polyploidy being important for the evolutionary success of angiosperms. Nevertheless, there are also some views that suggest polyploidy to be an evolutionary dead-end (Stebbins Jr., 1950; Wagner Jr., 1970; Mayrose *et al.*, 2011; Arrigo & Barker, 2012). This argument rests on evidence that show that neopolyploids diversify at lower rates and are more likely to become extinct in comparison to diploid lineages. However, as Soltis *et al*., (2014) pointed out, these studies focused on the aspects of polyploid extinction rate rather than the contribution of polyploids to evolutionary processes by generating genetic novelty. Although there is still much debate as to whether

polyploidy is beneficial or detrimental for a species' survival, the recurrence, ubiquity and multiple origins of polyploid species suggests that polyploidy is indeed an important evolutionary occurrence and not an evolutionary dead-end as previously thought.

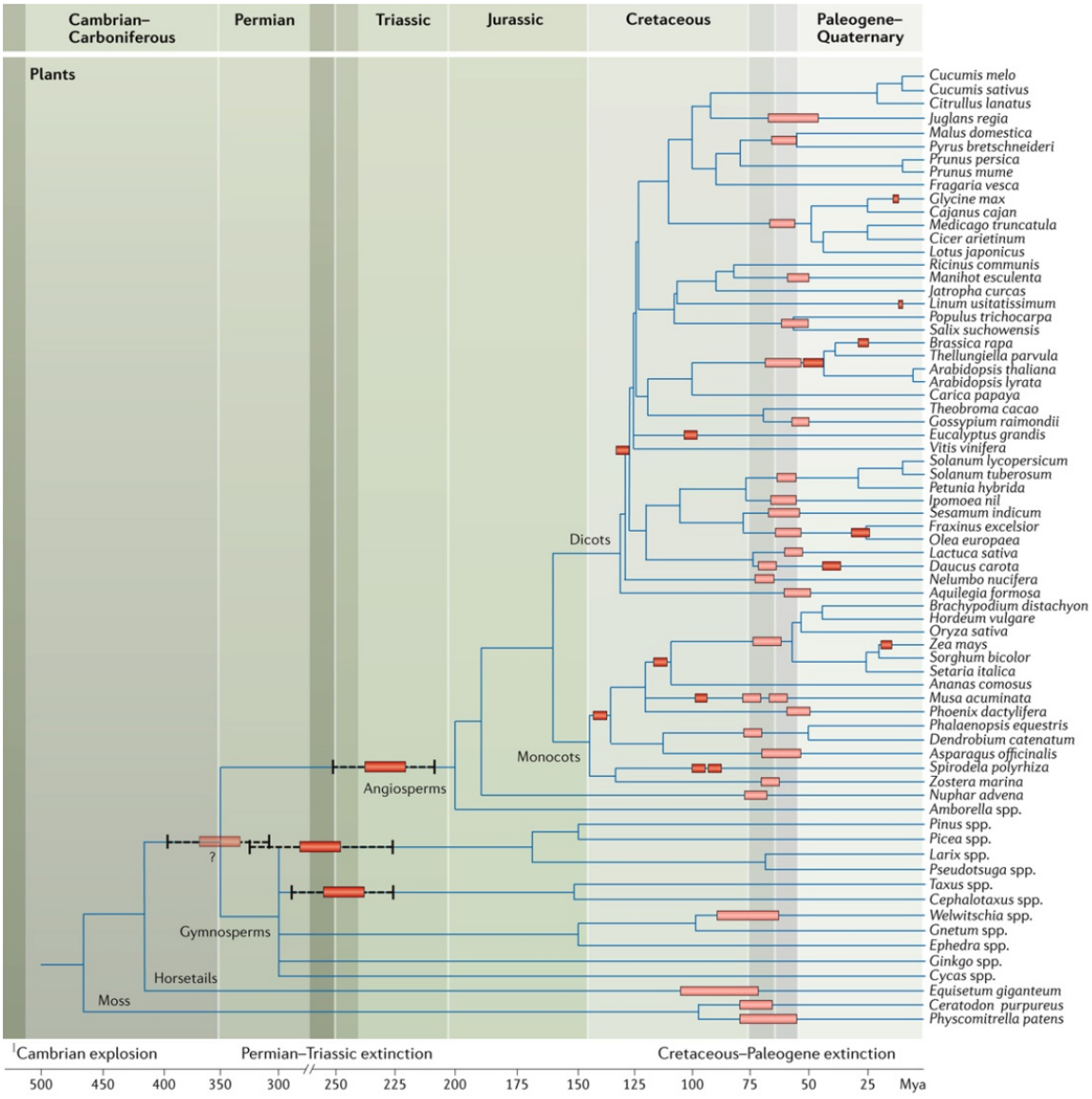


Figure 1.1 | Phylogenetic tree showing the evolutionary relationship between plant species in the context of polyploidy. Whole genome duplication events have been mapped and represented by red. Black dashes indicate uncertainty of the date of events. Mya = million years ago. Adapted from Van de Peer *et al.*, (2017).

1.3 Types of polyploidies

1.3.1 Autopolyploidy and allopolyploidy

Polyploids can originate through three different mechanisms: autopolyploidy, allopolyploidy, and triploid bridges (elaborated further in section 1.5.4). Autopolyploids are formed as a result of genome duplication within a species (Figure 1.2), e.g. potato (Stupar *et al.*, 2007), sugarcane (Zhang *et al.*, 2018), or banana (Heslop-Harrison & Schwarzacher, 2007; Amah *et al.*, 2019). In this case, all homologous chromosomes are identical. Allopolyploids on the other hand are formed by the hybridization between two (or more) species in either (i) single step : via the hybridization of unreduced male and female gametes from diploid species (Figure 1.2) or via hybridization between different autopolyploid species; (ii) two-steps : inter-species hybridization between diploid species and subsequent somatic genome doubling (Comai, 2005) (Figure 1.2). Some examples of allopolyploids include bread wheat (Haider, 2013), oat (Ansari & Thomas, 1983), coffee (Clarindo & Carvalho, 2008), and tobacco (Leitch *et al.*, 2008).

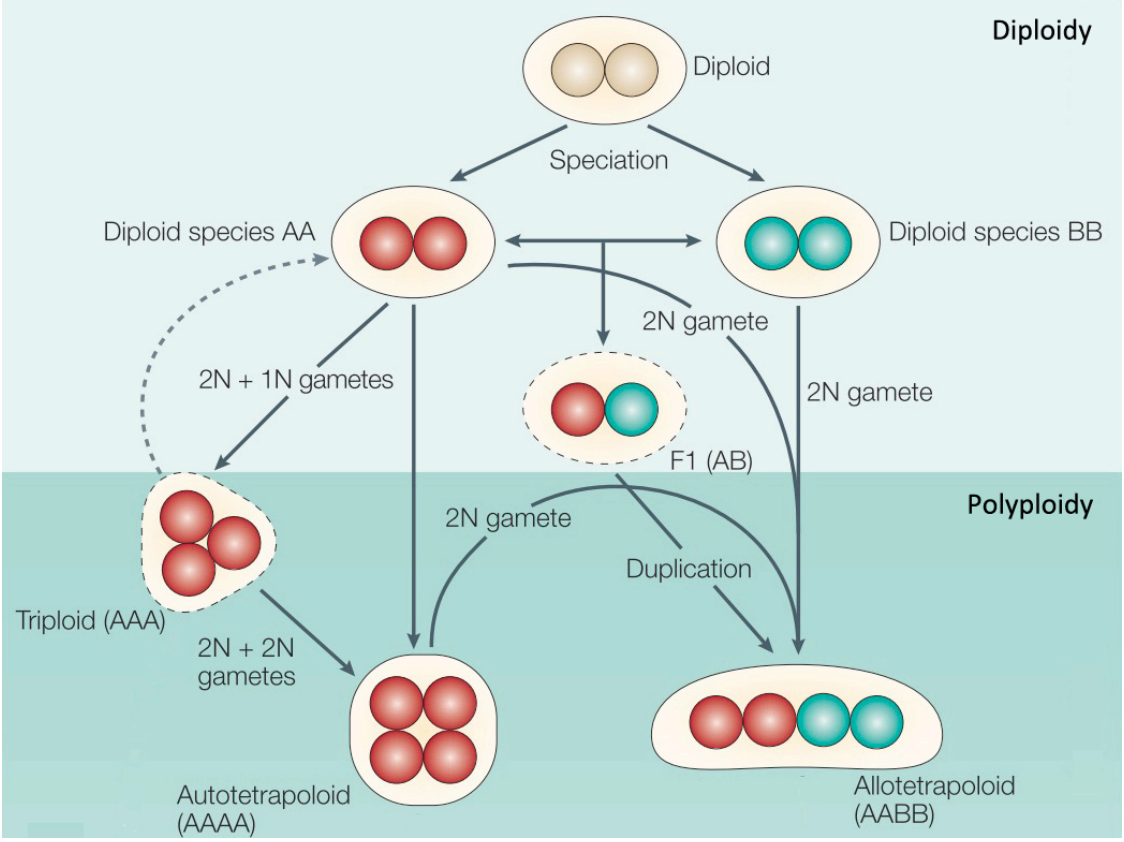


Figure 1.2 | Pathways to autotetraploid and allotetraploid formation. For each ploidy form, the haploid genome is represented by a coloured circle or oval inside the beige-filled nuclear shape. Circles or ovals of different colours represent diverged genomes. Highly unstable ploidy forms have dashed nuclear contours. A and B represent genome types and N is the gametic chromosome number. Adapted from Comai (2005).

Although binary in definition, in nature, autopolyploids and allopolyploids represent the extremities of a spectrum of genetic and taxonomic possibilities (Stebbins, 1971; Ramsey & Schemske, 1998, 2002; Bennett, 2004; Madlung, 2013; Doyle & Sherman-Broyles, 2017) (Figure 1.3). Depending on how closely related the parent species contributing to the polyploid are, a gradation of polyploids from true autopolyploids, to hybrid autopolyploids, to segmental polyploids, and allopolyploids can be formed (Stebbins Jr., 1950; Barker *et al.*, 2016). The classification of species into autopolyploids and allopolyploids can be challenging (Parisod *et al.*, 2010) as there exists a "grey zone" (De Queiroz, 2007) due to the subjectivity in assignment of taxonomic nomenclature to closely related species. This results in varying taxonomic schemes that make the concomitant categorisation of a polyploid species into autopolyploid or allopolyploid conflicting (Figure 1.3). For example, Barker (2016) and Doyle and Sherman-Broyles (2017) found discrepancies between their estimates of the number of autopolyploids and allopolyploids in *Glycine* species owing to under-curated databases, evolving taxonomy and ambiguous nomenclature of species.

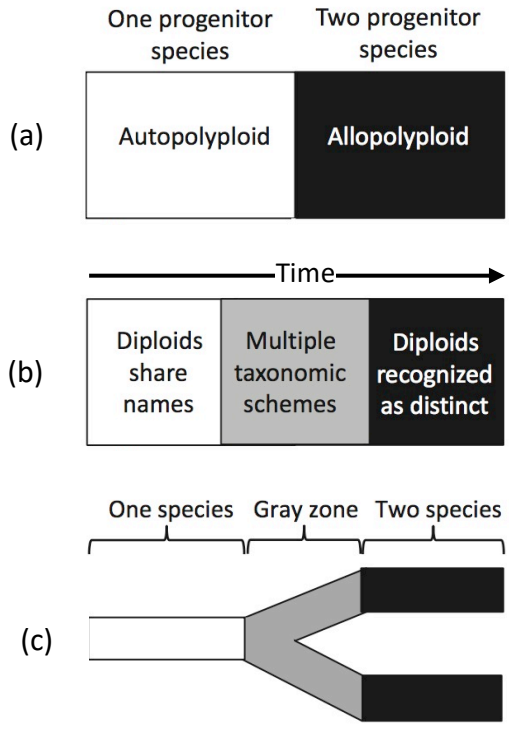


Figure 1.3 | Challenges when categorising a polyploid to either autopolyploids or allopolyploids (a) simple 'black and white' taxonomic definition of autopolyploids vs allopolyploids; (b-c) 'Grey-zone' in species classification due to (b) a transitionary phase between a single species (white rectangle) and two sister species (black rectangle) during diploid speciation, or (c) varying taxonomic criteria of species assignment

wherein it may be regarded as the same species or two different species depending on the subjectivity of the taxonomist. Thus, the same polyploid could be classified as an autopolyploid or an allopolyploid depending on whether two closely related progenitor species are considered the same or distinct. Adapted from Doyle & Sherman-Broyles (2017).

1.3.2 Paleopolyploidy and neopolyploidy

Polyploidy can also be classified temporally based on the time elapsed since genome duplication. Polyploids that occur as a result of ancient genome duplications several million years ago are known as paleopolyploids (Blanc & Wolfe, 2004b; Paterson, 2008). Given that all extant seed plants, irrespective of their current ploidy level are thought to have a polyploid ancestor (Jiao *et al.*, 2011), all plant species, including diploids, are theoretically paleopolyploids. Stable polyploid lineages of ancient origin exhibit the cumulative effects of genome duplication and natural selection. Because paleopolyploid genomes undergo significant genomic changes over evolutionary time, they are often not retained as an identical duplicated copy of the progenitor genome (even in autopolyploids) as discussed later in section 1.4.

To tease apart evolutionary effects from duplication effects, the study of neopolyploids i.e. newly formed polyploids has gained interest particularly over the past few decades. The development of synthetic polyploid lines from diploid lines has been the basis of many comparative studies aimed at understanding changes in polyploids relative to diploids on account of increased ploidy *per se,* and not due to the confounding effects of evolution. Comparative studies using synthetically formed autopolyploids (neoautopolyploids) are the focus of this thesis and will been discussed further in section 1.7.

1.3.3 Endopolyploidy

Endopolyploidy is different to the types of polyploids previously described, in that it occurs at a cellular level within certain tissues rather than at a whole-organism level. Thus, while we can have autopolyploid or neopolyploid plants for example, endopolyploid plants do not exist, but several plant species exhibit endopolyploidy in their cells or tissues (Nagl, 1976; Galbraith *et al.*, 1991; Joubès & Chevalier, 2000; Barow, 2006). Endopolyploid cells are commonly observed in leaves and trichomes of

Arabidopsis thalina (Roeder *et al.*, 2010; Bramsiepe *et al.*, 2010), roots of *Zea mays* (Li *et al.*, 2019), and stems of *Mesembryanthemum crystallinum* (Barkla *et al.*, 2018). In plants, endopolyploid cells usually arise through endoreduplication - a specialized mode of the cell cycle wherein cells enter the cell cycle, replicate their DNA during the S-phase but do not proceed to mitosis (cell division). This can occur for one or more rounds (called endocycles) giving rise to cells with doubled (one endocycle), quadrupled (two endocycles), or 2^n DNA content (where 'n' is the number of endocycles) (Edgar & Orr-Weaver, 2001; Leitch & Dodsworth, 2017). The purpose of endopolyploidy is not entirely clear, and they often have different distribution and roles in different species, but endopolyploidy has often been hypothesised to control the plasticity of organ size under stress (Gegas *et al.*, 2014), increase metabolic activity during seed and fruit development (Shu *et al.*, 2018), and promote root growth and structure formation for water metabolism (Barow, 2006; Bhosale *et al.*, 2018; Li *et al.*, 2019). Endoployploidy offers a good system to investigate the effect of higher DNA content in specific cell types, and how these could eventually impact the whole organism. While a fascinating topic, it will not be considered further in the context of this thesis.

1.4 Mechanisms shaping polyploid lineages

Genome duplication gives rise to gene redundancy and additional gene copies are rarely retained in their duplicated state. Due to sudden increase in the number of chromosomes, genome instability (a.k.a "genome shock") is seen for several generations after whole genome duplication, causing significant genomic reorganisation (Madlung *et al.*, 2005) that shape polyploid lineages. Discussed below are the processes that occur following genome duplication that act in isolation or in conjunction to ultimately determine the fate of duplicated genes over time.

1.4.1 Genomic rearrangements and genome downsizing

Immediately after whole genome duplication, newly formed polyploids are prone to mitotic and meiotic aberrations owing to the presence of additional sets of chromosomes. This is especially observed in allopolyploids, where the chromosome sets originate from two different genetic backgrounds, thus frequently causing non-homologous recombination during meiosis that gives rise to extensive genome rearrangements (Jenczewski *et al.*, 2003; Chen & Ni, 2006; Nicolas *et al.*, 2012; De Storme & Mason, 2014). Non-homologous recombination between distal regions (EET –

end to end translocations), or between distal and peri-centromeric regions (like NCI – Nested Chromosome Insertions, and Robertsonian translocations) often leads to gene loss and genome downsizing (Leitch & Bennett, 2004; Mandáková & Lysak, 2018). Genome restructuring following genome duplication has been reported in allopolyploid *Brassica* (Pires *et al.*, 2004)*, Nicotiana* (Lim *et al.*, 2007) and *Tragopogon* (Chester *et al.*, 2012), and to a lesser extent in some autopolyploid genera like *Saccharum* (Zhang *et al.*, 2018), *Paspalum* (Martelotto *et al.*, 2007) and *Arabidopsis* (Weiss & Maluszynska, 2000; Santos *et al.*, 2003).

1.4.2 Neofunctionalization

Following whole genome duplication, duplicate gene copies can gain novel functions over time – a process known as neofunctionalization (Figure 1.4). Dun *et al*., (2014) reported that in *Brassica napus* (an allotetraploid derived from two diploid species - *Brassica rapa* and *Brassica oleracea)*, neofunctionalization of the *BnaC9.Tic40* gene occurred wherein it gained a novel function related to male fertility while retaining its ancestral function to translocate proteins across the chloroplast inner membrane. Liu and Adams (2010) also showed evidence of neofunctionalizaion in the Brassicaceae family, however, here, one homolog *BSK1 (Brassinosteroid Kinase1*) retained its ancestral function and *SSP (SHORT SUSPENSOR*) lost its ancestral function and gained a new function. Differential neofunctionalization was seen between leaf types in *Zea mays*, wherein more neofunctionalization of genes was observed in foliar leaves as compared to husk leaves (Hughes *et al.*, 2014).

1.4.3 Subfunctionalization

Sometimes, genes do not gain novel functions but instead divide its function between the two homologs (Figure 1.4). This is known as subfunctionalization, following which both homologs of the gene pair need to be expressed in order to perform the ancestral function. A classic example of subfunctionalization is the *AGAMOUS* (*AG*)-like MADS-box gene in *Zea mays* where the ancestral function of *AG* gene is divided between the two homologs *ZMM2* and *ZAG1* (Mena *et al.*, 1996). In a recent study, D'Amelia *et al*., (2018) proposed that following duplication, homologs of the *R2R3 MYB* gene had subfunctionalized to segregate its functions of anthocyanin production (*AN1*) and cold stress response (*AN2*) in the potato species *Solanum commersonii*. Subfunctionalization

has also been proposed to be a transitionary phase between duplicated genes and neofunctionalization (Rastogi & Liberles, 2005).

1.4.4 Fractionation

The vast majority of duplicated genes are lost (Lynch & Conery, 2000), either through mutational loss of function of one of the homologous copies, or excision/deletion of chromosomal segments, collectively known as fractionation (Figure 1.4). Woodhouse *et al*., (2010) showed that duplicated genes in maize are predominantly lost, and that gene loss is due to deletion (not translocation) of chromosome fragments. Fractionation has also been reported in allohexaploid *Brassica rapa* (Tang *et al.*, 2012; Subramaniam *et al.*, 2013). Furthermore, several other studies have shown that gene fractionation is not stochastic and that certain gene categories like those that code for regulatory, ribosome or proteasome proteins, components involved in networks or signal transduction, and transcription factors are more resistant to fractionation (Thomas *et al.*, 2006; Freeling, 2009; Freeling *et al.*, 2015). On the other hand, some gene categories are more susceptible to fractionation and reverting to their single gene copy state like those involved in DNA recombination, replication and repair (De Smet *et al.*, 2013).

1.4.5 Epigenetic modification

In addition to DNA sequence changes on account of genome rearrangement, neo-functionalization, sub-functionalization, and/or fractionation, genome duplication may also trigger epigenetic changes i.e. changes in DNA methylation, histone modification, and chromatin remodelling (Chen, 2007). Epigenetic remodeling has been documented in allopolyploids of *Arabidopsis* (Lee & Chen, 2001; Madlung *et al.*, 2002; Tian *et al.*, 2014), *Brassica* (Lukens *et al.*, 2006; Xu *et al.*, 2009), and *Spartina* (Salmon *et al.*, 2005; Parisod *et al.*, 2009) among several others. There are relatively fewer studies in autopolyploids and they appear to elicit limited changes in epigenetic regulation in comparison. Studies in synthetic *Solanum* allopolyploids and autopolyploids showed that there was more DNA methylation in allopolyploids than in autopolyploids (Marfil *et al.*, 2018). Epigenetic modifications regulate gene expression by activating or silencing genes (Gent *et al.*, 2013; Song & Chen, 2015) and are thought to be involved with increasing diversity and plasticity through phenotypic novelty, thus facilitating

adaptation and establishment of stable polyploid lineages (Iwasaki & Paszkowski, 2014; Vogt, 2017; Thiebaut *et al.*, 2019).

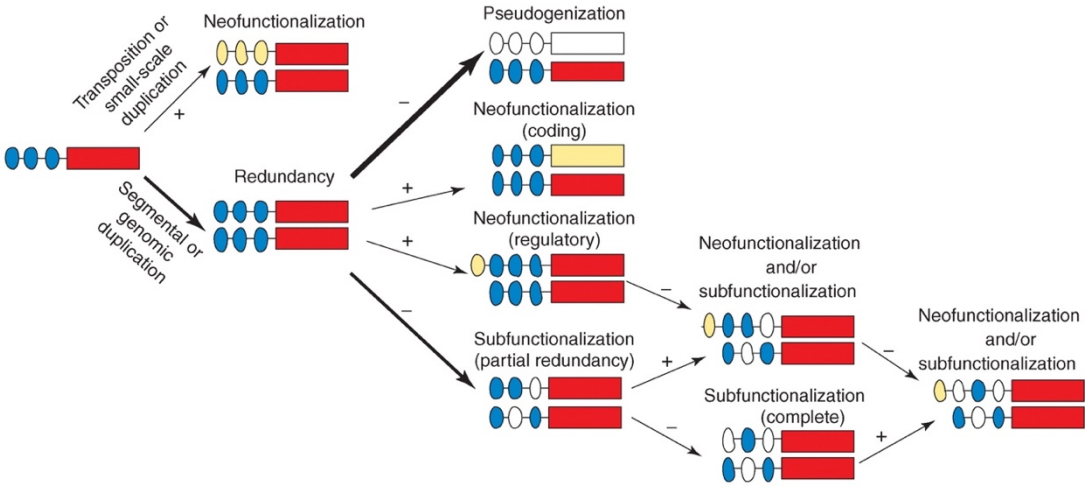


Figure 1.4 | Possible fates of duplicate gene copies following whole genome duplication. Redundant gene copies often lose their function (pseudogenization or fractionation), gain new functions (neofunctionalization), or divide their functions (sunfunctionalization). Ovals = regulatory subfunction (blue), gain of regulatory subfunction (yellow), loss of regulatory subfunction (white); rectangles = coding function (red), gain of coding function (yellow), loss of coding function (white); '+' represents a gain-of-function mutation and '-' represents a loss-of-function mutation. Adapted from Moore and Purugganan (2005).

1.5 Immediate changes in newly formed polyploids

1.5.1 Change in cellular homeostasis

Besides genetic and epigenetic changes that occur following genome doubling, changes are also observed at a cellular level (also known as nucleotypic changes). Doubling of the nuclear contents in the nuclei of cells i.e. increased genome size concomitantly increases the cell size. Arithmetically, increasing the cell size leads to a rapid increase in cell volume i.e cytoplasm ($V = r^3$; where 'r' is the radius of the cell) but translates to a $6r^2$ increase in surface area i.e membranes of the cell, nucleus, vacuole, and vesicles (Kondorosi *et al.*, 2000). This change in volume to surface area ratio is hypothesized to alter the ratio of interacting components, thus affecting cellular

homeostasis. This changes the relative concentration of interacting molecules and alters the rate of signal transduction, metabolism, and gene expression within the cell (Doyle & Coate, 2019) which in turn may alter plant phenotype and gene expression.

1.5.2 Mitotic instability

Wright *et al.,* (2009) studied natural and artificially created *Arabidopsis suecica* (allopolyploid) lines and demonstrated that there was frequent mitotic instability leading to somatic aneuploidy in natural *Arabidopsis suecica* lines, and either no or low-level somatic aneuploidy in artificial *Arabidopsis suecica* lines. Corneillie *et al.,* (2019) showed that increasing ploidy led to delayed growth and development in tetraploids, hexaploids and octoploids of *Arabidopsis thaliana*. This is hypothesized to be linked to slower cell division rates with increasing ploidy due to increased demand on resources and replication machinery (Tsukaya, 2008; Corneillie *et al.*, 2019).

1.5.3 Meiotic challenges

Meiotic challenges occur during pairing and segregation as there are more than one homologous chromosome to pair with. This often results in multivalent pairing during metaphase leading to the mis-segregation of chromosomes in anaphase (Yant & Bomblies, 2015). If odd number of chromosome sets are present (e.g. triploids or pentaploids), it is impossible to distribute them evenly and the random segregation of chromosomes mostly produces aneuploids (genomes with incomplete chromosome sets) (Henry *et al.*, 2005).

1.5.4 Barriers or bridges to gene flow

Polyploid lineages that establish and survive, often diverge to form new species as they tend to be reproductively isolated from their progenitor diploids (Arrigo & Barker, 2012) due to the production of either sterile (Ramsey & Schemske, 2002) or inviable (Scott *et al.*, 1998; Köhler *et al.*, 2010; Stoute *et al.*, 2012) F1 progeny. Diploid x tetraploid crosses can generate triploids which act as a “triploid block” to gene flow (Köhler *et al.*, 2010) that reproductively isolates ploidal levels and promotes speciation.

If triploid offsprings from interploidy crosses are fertile, they can either self-fertilize or back-cross with either parent to form an array of progeny ranging from diploids to aneuploids to tetraploids (Bergström, 1938, 1940; Burton & Husband, 2001;

Henry *et al.*, 2005), thus forming a "triploid bridge" that facilitates gene flow between varying ploidy levels (Ramsey & Schemske, 1998; Husband, 2004).

1.6 Background Summary

In summary, polyploidy is common in angiosperms is common, and the additional genetic raw material available in polyploids provides opportunities for the formation of novel allele combinations or gene redundancy that allow for genomic plasticity and consequently phenotypic variation that facilitates improved adaptation to new environments (Gaeta *et al.*, 2007; Leitch & Leitch, 2008; Doyle *et al.*, 2008; Chester *et al.*, 2012; Roulin *et al.*, 2013; Hao *et al.*, 2013). Polyploidy also facilitates the formation of new species by adaptive radiation (Wood *et al.*, 2009; Shimizu-Inatsugi *et al.*, 2017; Alix *et al.*, 2017).

While there are multiple mechanisms that shape polyploid lineages, they often overlap, interact, and/or act in conjunction, thus it may often be difficult to determine which genes are preserved, transformed or lost by specific processes. Genomic changes that are tolerated give rise to stable lineages and some of these plants persist as polyploids. However, with all these factors in play, over time, duplicated genomes often accumulate DNA sequence changes to the extent that the genome no longer represents an exact duplicate copy of the original diploid genome but rather possesses a pseudo-diploid like state – a process known as diploidization (Cuñado *et al.*, 2005; Conant & Wolfe, 2008; Renny-Byfield *et al.*, 2013; Dodsworth *et al.*, 2016; Poggio & González, 2018). Thus, in diploidized lineages, duplicated genomes transition from being tetrasomic (i.e. having four alleles at a locus) to being disomic (having two alleles at each of two distinct loci).

Historically, the lack of established autopolyploid lineages was interpreted as autopolyploidy being a disadvantage and an evolutionary dead-end (Stebbins Jr., 1950). This view has predominantly changed over time as studies suggest polyploidy to be a major driver in angiosperm evolution (Madlung, 2013; Soltis *et al.*, 2014; Dodsworth *et al.*, 2016). With the diploidization of paleopolyploid lineages being common, it remains unclear whether extant polyploid species from ancient polyploid lineages have persisted by random chance or have been selected over evolutionary time because of their higher ploidy (Hunt *et al.*, 2011; Oswald & Nuismer, 2011; Čertner *et al.*, 2019).

To understand polyploidy comprehensively, it is important to understand what ploidy does to the plant lineages in which it occurs, and the advantages or disadvantages that it confers, both immediately and in the long term. While noteworthy research has gone into understanding the effects of ploidy on the success of paleoallopolyploid plant lineages, those aimed at understanding the more immediate effects, especially in neoautopolyploids are sparse (Spoelhof *et al.*, 2017). Therefore, to separate the effects of ploidy from the effects of the merger of different genomes (allopolyploidy) or the long-term effects of polyploidy and its interaction with natural selection (paleopolyploids), it is important to concentrate on newly formed autopolyploids (neoautopolyploids).

1.7 Thesis focus: Neoautopolyploidy

1.7.1 Methods of artificially inducing neoautopolyploidy

In nature, spontaneous doubling of chromosomes may occur, giving rise to natural neoautopolyploids, however, this is rare and often difficult to identify (Tayalé & Parisod, 2013). To understand neoautopolyploids better, artificial synthesis via the somatic doubling of the chromatin contents in cells of early embryonic plant tissue (Ramsey & Schemske, 1998; Ascough *et al.*, 2008) is preferred as the process can be tracked from formation of the polyploid through to the changes observed in each subsequent generation. Artificial synthesis is also reproducible and can be carried out simultaneously for a range of plant species and a number of replicates that allows for side-by-side comparisons that is difficult to obtain with natural neopolyploids.

While several methods to stimulate artificial genome doubling like ionizing radiation (De Nettancourt *et al.*, 1971), temperature shock (De Storme *et al.*, 2012), and nitrous oxide gas (Kitamura *et al.*, 2009) exist, chemical treatment remains the common method of choice due to its relative ease of application and effectiveness in inducing genome doubling. Chemical treatment with anti-mitotic agents alter the plants' cell cycle by interfering with spindle assembly at metaphase. This prevents the cell from transitioning to anaphase as chromosomes cannot separate, resulting in cells with doubled chromosome numbers (Planchais *et al.*, 2000; Dhooghe *et al.*, 2011). Chemicals like colchicine, trifluralin and oryzalin are mitotic inhibitors that have been used to generate artificially induced 'neo-polyploids' since the late 1960's (Semeniuk & Arisumi, 1968; Lignowski & Scott, 1972), however, their use as a tool in understanding

mechanisms underlying neopolyploidy has been fairly recent (Husband *et al.*, 2008; Hegarty *et al.*, 2013). The use of chemicals to create synthetic autopolyploids is critical to understand the effect of genome duplication in the early stages.

1.7.2 Methods of determining ploidy of potential polyploids

Chromosome doubling using chemical agents can often be a 'fickle' process, as the right choice of chemical, its concentration and method of application are vital for successful polyploidization. Even after standardizing induction protocols, it does not guarantee chromosome doubling. Thus, progeny arising from chemically treated plants need to be tested to ascertain their ploidy in order to (i) know if the polyploidization has been successful (ii) determine the ploidy level of the plant (i.e. tetraploid, octoploid etc.). Furthermore, chemical induction of polyploidy is usually achieved by the application of the chemical solution to young, actively growing shoot apical meristem tissue. This often results in the formation of chimeric cell populations that give rise to mixoploid plants containing tissues of varying ploidy levels (Kamwean *et al.*, 2017). Thus, it is important that seeds obtained from treated plants are selfed for at least one generation before ploidy assessment, in order to ensure that pure lines are established.

The most reliable way to determine chromosome number is by microscopically counting chromosomes in metaphase spreads (also known as a chromosome squash) of actively dividing cells that have been stained with nuclear dyes (Maluszynska, 2003). This gives the exact chromosome number and allows for the detection of aneuploidy (if any). However, the process can be time consuming and labour intensive, especially when large number of lines are to be assessed. Moreover, for plants that have small genomes (e.g. *A. thaliana*), visualising and distinguishing between chromosome numbers is particularly difficult. Given the challenges that chromosome counts pose, the estimation of DNA content using flow cytometry is often favoured and considered to be a reasonable alternative. Using flow cytometry, thousands of cells can be rapidly and accurately examined for each sample, and the DNA content in the nucleus can be determined based on the fluorescence intensity detected by the flow cytometer relative to those of the internal controls of known diploid DNA values. DNA values obtained are then compared to the DNA content values of diploid lines from the same species and divided to determine ploidy (Ochatt, 2008). Flow cytometry is a relatively quick and effective method to determine ploidy, however, as it does not estimate exact

chromosome numbers, it may not efficiently detect aneuploidies. Nevertheless, it remains a widely accepted tool for ploidy verification.

1.7.3 Effects of neoautopolyploidy

In polyploids where DNA sequence changes occur either due to genome merger (in allopolyploids), or evolutionary processes restructuring genomes over time (in paleopolyploids), a change in gene expression, phenotype and reproduction can be expected owing to novel gene combinations or altered gene functions. However, in neoautopolyploids, given that minimal (or no) sequence changes would be expected immediately after artificial autopolyploidization, any changes observed should be brought about by bulk increase in DNA amounts only (i.e. nucleotype changes). Such genome scaling effects or 'gigas effects' should, in theory, be easy to elucidate as the changes they bring about should be the same irrespective of the plant genotype or species in question. Nucleotype-related effects have been documented in several neoautotetraploid species like *Nicotiana attenuata* and *Nicotiana obtusifolia* that displayed increased stomatal size, seed weight and dry biomass weight (Anssour *et al.*, 2009); and increased cell size in *Citrus limonia* (Allario *et al.*, 2011) and *Zea mays* (Yao *et al.*, 2011). In addition to nucleotypic effects, phenotypic variation, variation in gene expression and variation in interploidy hybridization responses have been observed (explored in detail in chapters 2, 3 and 4 respectively), suggesting that other mechanisms, unrelated to cell size increase also contribute to polyploidy responses. Neoautopolypoid studies that specifically address this variation in response to increased ploidy are limited, and further exploration of these responses can prove vital to the complete understanding of mechanisms that govern responses to genome doubling *per se* immediately after a whole genome duplication event.

1.7.4 Genetic variation in neoautopolyploids

Plants show widespread genetic variation. As highlighted by Alonso-Blanco and Koornneef (2000), studying functional processes in the light of natural variation can be an excellent tool to unravel the genetic basis underlying complex traits (Borevitz & Nordborg, 2003). Although the use of natural variation as a tool to explain phenotypic traits, adaptive processes and crop domestication has been increasing (Doebley *et al.*, 2006; Mitchell-Olds *et al.*, 2007; Alonso-Blanco *et al.*, 2009; Henderson & Salt, 2017;

Coolen *et al.*, 2019; Duruflé *et al.*, 2019), their use in the context of understanding polyploid processes has been sparse.

Thus far, most neoautopolyploids studies have been restricted to one or two genetic backgrounds (i.e. genotypes, ecotypes, or accessions) within a species, which provides limited insight as to how increased ploidy affects the species beyond the genotype being studied. In order to extrapolate findings to the species-level, experiments need to be conducted in multiple genotypes. Understanding why certain genotypes exhibit different responses to increased ploidy can help underpin the genetic basis of ploidy responses. As most complex traits, and perhaps ploidy responses, are polygenic, using natural variants to study them provides an advantage over loss-of-function mutation studies or other traditional gene-by-gene approaches as the latter may exhibit lethal phenotypes or gene redundancy, making the interpretation of their specific gene function problematic (Borevitz & Nordborg, 2003). Natural variation may not only assist in elucidating functional characterization of genes involved in ploidy responses but may also help in understanding gene interactions and networks that work in collaboration to regulate gene expression, phenotypic, and reproductive traits.

1.8 Arabidopsis thaliana as a model for ploidy studies

Arabidopsis thaliana has been an organism of common choice for plant studies as it has a short life cycle, a relatively small and completely sequenced genome (Bevan & Walsh, 2005). The easy availability of various types of *A. thaliana* lines from the two stock centers (NASC and ABRC) as well as the accessibility of dedicated databases and online tools makes it a model system that is much sought after (Provart *et al.*, 2016). *A. thaliana* is a self-fertilizing diploid, with a wide geographical distribution that has a large number of naturally diverging homozygous genotypes (also referred to as “accessions”) that provide valuable resources for understanding the molecular basis of various traits (Alonso-Blanco & Koornneef, 2000; Borevitz & Nordborg, 2003; Koornneef *et al.*, 2004; Weigel, 2012)

For neoautopolyploids studies in particular, *A. thaliana* serves as a good model due to the relative ease of developing neopolyploid lines artificially via treatment with colchicine (Santos *et al.*, 2003; Jeffrey Chen *et al.*, 2004; Yu *et al.*, 2009; Hegarty *et al.*, 2013). While *A. thaliana* is mostly diploid, there are a few related species that are polyploid, such as *Arabidopsis arenosa* and *Arabidopsis lyrata (*both of which also have

diploid lineages), and *Arabidopsis suecica* and *Arabidopsis kamchatica* (which only occur as tetraploids) (Bomblies & Madlung, 2014).

The short life cycle combined with in-depth genomic data makes it an ideal model system to investigate the effect of autopolyploidy in its initial stages. The natural existence of many homozygous genotypes allows the investigation of polyploid responses in various genetic backgrounds. A limited number studies have looked into natural genetic variation in response to increased ploidy in *A. thaliana* (but see Yu *et al.*, 2010; Monda *et al.*, 2016). Both studies used two genotypes and found notable differences in tetraploid gene expression and phenotypic responses between the tetraploid genotypes. Despite being a promising avenue, this topic has been vastly understudied in *A. thaliana,* and across other plant species as well, with only a handful of studies, all of which provide remarkable findings.

A. thaliana also serves as a valuable model to study the genetic basis interploidy hybridization barriers (Scott *et al.*, 1998; Henry *et al.*, 2005; Dilkes *et al.*, 2008; Kradolfer *et al.*, 2013; Schatlowski *et al.*, 2014a; Duszynska *et al.*, 2019). When diploids hybridize with tetraploids, they result in triploids of varying fertility. These triploids, if fertile, can back-cross with either parent, thus forming a "triploid bridge" to facilitate gene flow between varying ploidy levels (Ramsey & Schemske, 1998; Husband, 2004). If the triploid progeny are sterile, they form a "triploid block" to gene flow (Köhler *et al.*, 2010), thus reproductively isolating ploidal levels and promoting speciation. The mechanisms underlying triploid bridges and triploid blocks are complex and understanding them can have implications on our understanding of evolutionary processes and hybridization barriers.

1.9 Thesis aims

In the light of the literature explored in the previous sections, we found that polyploidy research in the area of autopolyploidy, and more specifically in the field of natural genetic variation in phenotypic, gene expression and reproductive responses to increased ploidy remain limited. In this thesis, we focus on neoautopolyploids to tease apart the effects of increased ploidy from other confounding effects. Using the anti-mitotic chemical agent colchicine, we treated diploid *A. thaliana* lines from multiple genotypes to generate colchicine-treated tetraploid and octoploid lines. Since the effects of colchicine treatment, independent of polyploidization, have not yet been fully understood, we used colchicine treated lines that did not undergo genome doubling as

diploid controls alongside untreated diploids. Using multiple genotypes of diploid and polyploid *A. thaliana* lines, we designed experiments to test whether the effects of neopolyploidy are genotype-specific or are independent of genotype.

In chapter 2, we studied the effects of tetraploidy on various phenotypic traits in seven genotypes of *A. thaliana*. We examined whether the phenotypic effects of increased ploidy were dependent or independent of genotype. Additionally, we also assessed whether colchicine had any hereditary or stochastic effects that distorted the interpretation of phenotypic data, and accounted for such effects in the analyses models to eliminate all confounding effects that interfered with the interpretation of ploidy effects on phenotype.

In chapter 3, we used a polyploidy series of diploids, tetraploids and octoploids for two genotypes of *A. thaliana* to study gene expression changes in response to increased ploidy. We examined whether there were any phenotypic and gene expression changes between the ploidy levels and if they were genotype dependent. As we used three ploidy levels, we were also able to address whether gene expression changes followed a linear pattern across the three ploidies. Finally, we used the gene expression data to explain the phenotypic variation observed in the polyploids.

In chapter 4, we used a special type of *A. thaliana* recombinant inbred line known as the MAGIC lines (Kover *et al.*, 2009) that are derived from multi-parent populations. To determine whether seed abortion (or triploid block) responses in paternal excess interploidy crosses were genotype-specific, we crossed the MAGIC lines to two paternal tetraploid genotypes to identify quantitative trait loci (QTL) and fine map seed abortion responses for each genotype.

Chapter 2 : **NATURAL GENETIC VARIATION IN PHENOTYPIC RESPONSES TO INCREASED PLOIDY IN *ARABIDOPSIS THALIANA*.**

Sarah Chordekar, Heather Phillips, Ilia J. Leitch, Jaume Pellicer and Paula X. Kover

Author Contributions :

SC : Performed the experiments, analysed the data, wrote the draft

HP : Performed the experiments

IJL, JP : Performed the flow cytometric analysis for ploidy testing

PXK : Supervised, reviewed the draft.

2.1 INTRODUCTION

Polyploidy is the presence of three or more homologous sets of chromosomes within the nucleus of an organism. It can occur as a result of chromosome multiplication within the same species (autopolyploidy) or from the hybridization of two different species (allopolyploidy) (Comai, 2005). Although polyploidy is not commonly tolerated in animals (except in certain groups of fish and amphibians), its occurrence in the plant kingdom is widespread (Albertin & Marullo, 2012; Soppa, 2014; Van de Peer *et al.*, 2017). Several studies suggest that polyploid plants may often exhibit advantageous phenotypes over their diploid counterparts, however, the reason for this pattern is not entirely understood. Understanding how ploidy affects plant traits can have several commercial applications in plant breeding (Sattler *et al.*, 2016), yield improvement (Serapiglia *et al.*, 2014; Eliášová & Münzbergová, 2014; Münzbergová & Skuhrovec, 2017), disease resistance (King *et al.*, 2012; Nellist *et al.*, 2019), and stress tolerance (Udall & Wendel, 2006; Ramsey, 2011; Allario *et al.*, 2013; Renny-Byfield & Wendel, 2014; Wei *et al.*, 2018).

Determining the effects of ploidy on plant phenotype in existing polyploid lineages is challenging due to the difficulty in distinguishing the effects of increased ploidy from hybridization effects (in allopolyploids) and the effects of natural selection (in paleopolyploids) (Husband *et al.*, 2008, 2016; Oswald & Nuismer, 2011). Thus, in order to isolate and study the effects of ploidy *per se*, a commonly used approach is to generate artificially induced synthetic autopolyploid lines also known as 'neoautopolyploids' (Husband *et al.*, 2008).

Neoautopolyploid studies in various species have shown that when compared to diploids, polyploids are more tolerant to biotic and abiotic stresses, for example, salinity and drought stress in *Arabidopsis thaliana* (Del-Pozo & Ramirez-Parra, 2014), drought stress in *Medicago sativa* (Zhang *et al.*, 2015), heat stress in *Dioscorea zingiberensis* (Zhang *et al.*, 2010a), and fungal infections in *Malus domestica* (Hias *et al.*, 2018). Although the altered phenotypes and improved tolerance of neoautopolyploids to various biotic and abiotic stresses have been widely recognized, the mechanisms contributing towards them are not well understood.

In neoautopolyploids especially, it is particularly interesting to understand how increasing the number of homologous chromosomes alters phenotypes, since the immediate aftereffect of duplication should only be an increase the number of gene copies, and not the repertoire of genes or possible gene interactions. A commonly proposed hypothesis is the 'gigas' effect or 'nucleotype' effect (Stebbins Jr., 1950; Levin, 1983) wherein the presence of additional DNA in the nucleus leads to an increase in cell volume, which directly affects cell and organ size (Kondorosi *et al.*, 2000; Knight & Beaulieu, 2008), as well as affects the development and physiology through changes in diffusion and transport of signals within cells (Ramsey & Schemske, 2002; Doyle & Coate, 2019). If this is the sole mechanism responsible for altered polyploid plant phenotype in comparison to diploids, then an increase in bulk DNA contents in plant cells should affect phenotypes of all plants in a similar manner, irrespective of the genotype of the plant (Bennett, 1971; Bennett & Riley, 1972; Símová & Herben, 2012; Doyle & Coate, 2019).

Very few studies have investigated the genotype-specific effects of increased ploidy on phenotype in newly formed autopolyploids. Riddle *et al.,* (2006) studied four inbred lines at three ploidy levels in *Zea mays* for 13 phenotypic traits and found that some traits showed a common response and some had genotype-specific responses to increased ploidy. Hias *et al.,* (2017) found genotype-specific changes in phenotype in two genotypes of *Malus domesticus* neotetraploids. Other studies have found that chromosomal rearrangements, epigenetic effects or alternative splicing that may occur in allopolyploids are biased to specific loci. If similar mechanisms operate in autopolyploids, then the genetic composition of the plant would affect genetic, epigenetic or RNA modifications and consequently alter gene expression that can modulate gene expression networks controlling quantitative traits. Thus, when studying the effects of ploidy on plant phenotype, an important starting point would be look at multiple genotypes to determine if there are any patterns of general (non-specific) phenotypic responses, and any genotype-dependent (specific) responses to increased ploidy. Understanding this may help unravel additional mechanisms underlying ploidy responses that are not passively regulated by bulk increase in DNA amounts (i.e. nucleotype), but rather caused due to more complex mechanisms regulated by the genetic composition of the plant.

Over the past decade, the vast majority of neoautopolyploid studies across various plant species have used colchicine to artificially induce genome duplication and then compare the neopolyploids (formed as a result of colchicine treatment) to progenitor diploids (that have never been exposed to colchicine) to explain ploidy effects (Yu *et al.*, 2009; Zhang *et al.*, 2010b; Meng *et al.*, 2011). Colchicine is a naturally occurring compound extracted from autumn crocus (*Colchicum autumnale*), which when applied to the apical meristem of diploid seedlings inhibits spindle formation and arrests the cell at metaphase by inducing microtubule depolymerization (Nebel, 1937; Davidson, 1961; Caperta *et al.*, 2006; Lu *et al.*, 2012). While the absence of microtubules prevent chromosome segregation leading to the production of tetraploid cells (Caperta *et al.*, 2006), it is possible that this mode of action could have other consequences, leading to other cytogenetic abnormalities in addition to the intended tetraploidization effect (Ramsey & Schemske, 2002). Thus, when comparing neopolyploids (colchicine-treated) to progenitor diploids (untreated) and ascribing any phenotypic changes recorded to ploidy, studies either assume that colchicine does not have any effects besides increasing ploidy, or any colchicine effects are non-hereditary, thereby overlooking potential side-effects the chemical may have. However, a few studies have shown that colchicine can have side-effects that alter plant phenotype independent of genome doubling (Husband *et al.*, 2008, 2016; Pignatta *et al.*, 2010) that last for at least two generations post treatment (Münzbergová, 2017).

Here, we used an experimental design that comprehensively isolated the effects of ploidy from the effects of hybridization, natural selection, and colchicine. To understand natural genetic variation in phenotypic responses to increased ploidy, we used multiple diploid and neoautotetraploid *Arabidopsis thaliana* lines from multiple genotypes. We tested whether the phenotypic effects of ploidy could be explained explicitly by the doubling of the genetic material by investigating if the phenotypic effects of neopolyploidization in autopolyploids were dependent or independent of genotype. We controlled for the effect of colchicine by including diploid and tetraploid lines that had both been exposed to colchicine. To control for any variable stochastic effects of colchicine, we generated multiple independently derived individuals of colchicine-exposed diploid and tetraploid lines for each genotype to account for stochastic variation within lines of the same genotype and ploidy level. By using plants

three generations after the colchicine treatment, we also addressed whether colchicine had hereditary effects that needed to be controlled. Our experimental design eliminated all confounding effects that interfered with the interpretation of ploidy effects, thus providing a more accurate understanding of the genotype-dependent and genotype-independent effects of ploidy on plant phenotype.

2.2 MATERIALS AND METHODS

2.2.1 Colchicine treatment of A. thaliana lines

Seven genotypes (natural genotypess) of *Arabidopsis thaliana* were used in this study (Table 2.1). Diploid seeds were obtained from the Arabidopsis Biological Resource Centre (ABRC, USA) and grown through self-seeds for a few generations in the Kover Lab. These lines originate from a wide geographic distribution and are a good source of natural genetic and phenotypic variation for studying ploidy response.

Table 2.1 | List of germplasm numbers and abbreviated names of the genotypes used in this study.

Genotype	ABRC Germplasm Number	Abbreviated Name
Columbia	CS6673	Col
Catania	CS6674	Ct
Kaunas	CS6762	Kn
Landsberg	CS20	Ler
Martuba	CS1380	Mt
Nossen	CS6805	No
Wurzburg	CS6897	Wu

To obtain new tetraploid lines, the shoot apical meristem of 50-80 two-week old seedlings of each genotype were treated with 10 µl 0.1% (w/v) colchicine (Sigma-Aldrich) and were grown to maturity in individual 2" pots containing compost (F2+ S compost from Levington Seed and Modular Compost, Scotts Company, UK). Treated plants were designated as generation G_0. Selfed seeds from each G_0 plant that survived the colchicine treatment and produced seeds were collected in individual envelopes for each plant. Selfed seeds (one seed from each G_0 plant) were grown in separate pots to generate G_1 plants.

2.2.2 Ploidy determination of colchicine treated lines

Nuclear DNA content of G_1 plants was assessed by flow cytometry at the Jodrell Laboratory (RBG, Kew)[1] using propidium iodide (PI) as a nuclear stain as outlined in Pellicer *et al.*, (2012) with some modifications. Briefly, ~1cm^2 leaf tissue from the target sample was co-chopped with *Oryza sativa* 'IR36' (internal reference standard; 1C = 0.5 pg) leaves using a sharp fresh razor blade in a petri dish containing 1mL ice-cold LB01 buffer (Doležel *et al.*, 1989). The nuclei suspension was passed through a 30 µm nylon mesh and stained with 100µL PI. The relative fluorescence was estimated using a flow cytometer (Partec GmbH, Münster, Germany). A minimum of 3,000 nuclei were analysed per sample and the histograms obtained were analysed. Nuclear DNA content of the sample was calculated using the formula:

$$\textbf{Sample 2C nuclear DNA content [in pg]} = \frac{\textbf{Mean of sample peak} \times \textbf{\textit{Oryza sativa}} \textbf{ 2C DNA content [in pg]}}{\textbf{Mean of \textit{Oryza sativa} peak}}$$

Since endoreduplication occurs in *Arabidopsis thaliana* leaves (Sugimoto-Shirasu *et al.*, 2002), multiple peaks were observed corresponding to 2C, 4C, 8C and 16C nuclei. Thus, only the mean value of the first sample peak (2C = 2x for diploids and 2C=4x for tetraploids) was used to calculate the value of the mean sample peak.

2.2.3 Genotyping of ploidy-tested lines

To ensure that there were no cultivation errors, each ploidy-tested line was genotyped using a set of five primer pairs – one for each chromosome (Supplementary Data S2.1) to ascertain that it was the correct genotype. We designed the primers using data on large indels (insertions and deletions that were ~100-200bp long) mapped by Gan *et al.*, (2011) to unequivocally identify each of the seven genotypes by producing an electrophoresis gel banding pattern unique to each genotype. Each 25µL PCR reaction contained 2µL DNA (~20ng/µL), 1.5µL forward primer (10µM), 1.5µL reverse primer (10µM), 12.5µL 2x DreamTaq Green PCR Master Mix (Thermo Fisher Scientific) and 7.5µL nuclease free water. PCR amplification of 40 cycles was carried out (92°C for 15s, 50°C for 30s, 72°C for 90s) with an initial denaturation step at 92°C for 2mins and a

[1] Flow cytometry analysis was attempted in-house by the candidate on a BD FACS Canto flow cytometry system. However, due to limited technical expertise and/or the equipment not being optimised for the analysis of small nuclei, the attempts were unsuccessful. Ploidy determination was therefore outsourced for efficiency, accuracy and reliability of interpretation of the flow cytometry histograms. Thus, this part of the experiment has not been performed by the candidate.

final extension step at 72°C for 5mins. The PCR products were run on a 1% agarose gel, stained with Ethidium bromide and viewed under UV light in a gel doc system. The genotype was determined based on the size of products obtained with each primer pair. The expected PCR product length for each genotype and primer pair is available in Supplementary Data 7.1.

2.2.4 Experimental Design

For each genotype, we retained selfed seeds from at least three G_1 plants confirmed to be the correct genotype and tetraploid after the colchicine treatment ($4x_{\text{-colchicine}}$) and three G_1 plants that remained diploid after the colchicine treatment ($2x_{\text{-colchicine}}$). Since colchicine treatment did not always generate tetraploids, we retained the colchicine treated lines that were diploid and used them as controls alongside untreated diploid lines ($2x_{\text{-untreated}}$). This allowed us to determine if colchicine treatment itself had any effect on the plant phenotype that was unrelated to ploidy. The G_2 tetraploid ($4x_{\text{-colchicine}}$) and diploid ($2x_{\text{-colchicine}}$) seeds obtained were further selfed for one more generation to produce G_3 seeds (as illustrated in Figure 2.1) that were used as lines for the phenotypic characterisation. In total, we used 48 lines (22 $4x_{\text{-colchicine}}$ lines, 19 $2x_{\text{-colchicine}}$ lines, and 7 $2x_{\text{-untreated}}$ lines) spanning across seven genotypes as outlined in Table 2.2. Lines that originated from independent events of colchicine treatment that belonged to the same genotype and of the same ploidy were distinguished by letters A, B, C and D (e.g. the three colchicine treated tetraploid Columbia lines are named Col 4x – A, Col 4x – B and Col 4x – C).

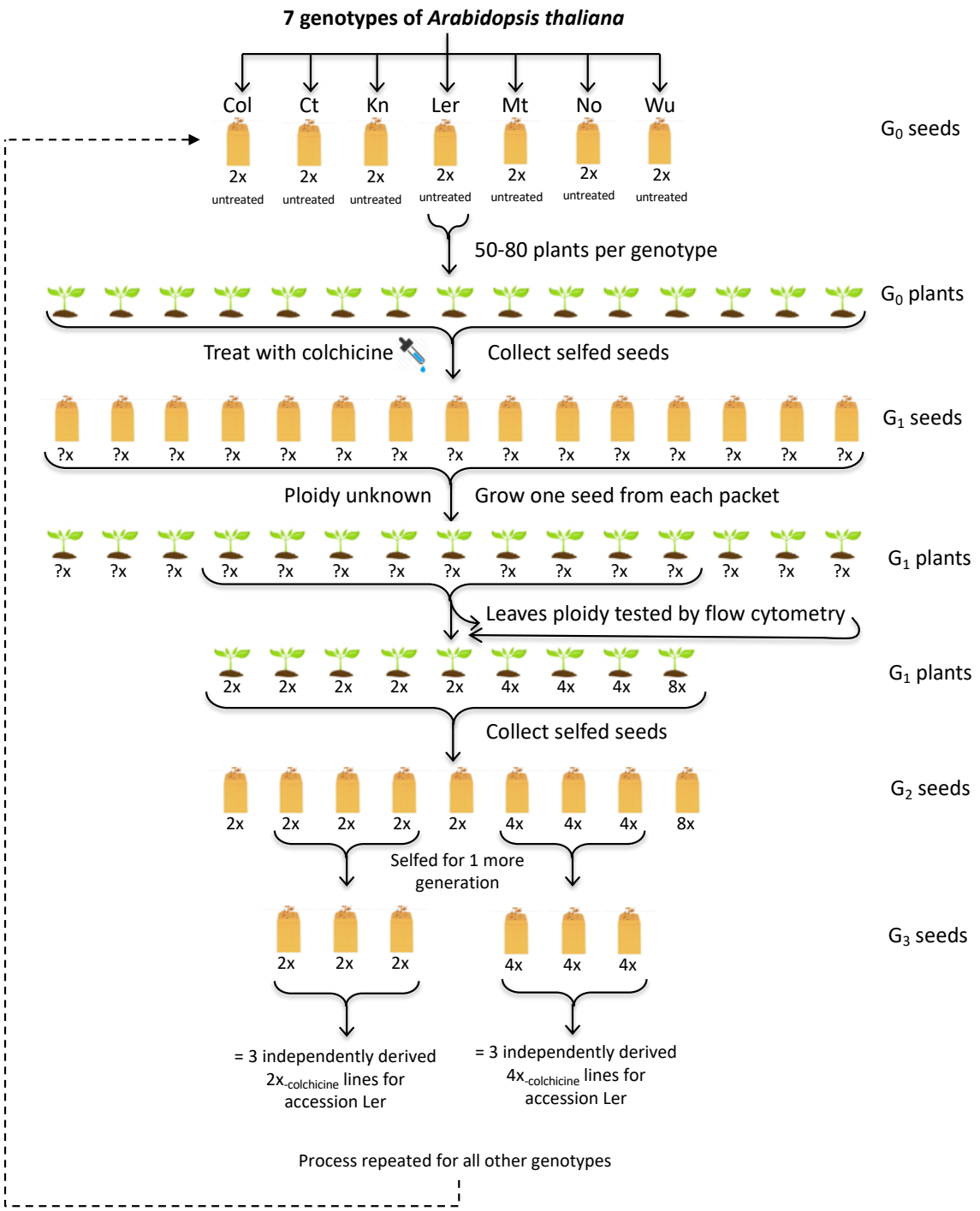


Figure 2.1 | Schematic representation of the generation of colchicine treated diploid and tetraploid lines. G_0 = base generation; G_1 = first generation after selfing; G_2 = second generation after selfing; G_3 = third generation after selfing. For each genotype, 50-80 plants were grown, treated with 0.1% (w/v) colchicine solution and selfed for one generation. Leaves from a subset of G_1 plants were ploidy tested, and their seeds collected. G_2 seeds were selfed for another generation and three[2] independently derived diploid and tetraploid lines for each genotype were used in this experiment.

[2] For some genotypes, we were unable to obtain three lines for each ploidy level (see Table 2.2).

Table 2.2 | A summary of the number and types of independently derived lines for each genotype used in this experiment. One $2x_{-untreated}$ line, three $2x_{-colchicine}$ lines and three $4x_{-colchicine}$ lines were used for four of the seven genotypes. For genotypes Ct and No we had two $2x_{-colchicine}$ lines and four $4x_{-colchicine}$ lines, and for Wu we had two $4x_{-colchicine}$ lines instead of three lines for each. Thus, a total of 48 lines were phenotyped instead of the expected 49. Ten replicate plants were phenotyped for each line.

Genotype	$2x_{-untreated}$	$2x_{-colchicine}$	$4x_{-colchicine}$	Total lines
Col	1	3	3	7
Ct	1	2	4	7
Kn	1	3	3	7
Ler	1	3	3	7
Mt	1	3	3	7
No	1	2	4	7
Wu	1	3	2	6
Total lines				**= 48**
Total plants phenotyped (N)				**= 480**

2.2.5 Phenotypic Characterisation

For phenotypic characterisation, seeds from all lines were cold stratified in the dark at $4^{O}C$ for five days in Eppendorf tubes containing 1mL water in order to promote synchronous germination of the seeds after sowing. Five seeds were then sown into each of ten replicate pots per line, for a total of 480 pots (48 lines x 10 replicates) that were randomly arranged in 20 trays in a growth chamber (Percival AR-66L) set at $22^{O}C$,16-hour light/$20^{O}C$, 8-hour dark photoperiods. The pots were 2" in diameter and contained soil (F2 + S variety of compost from Levington® Seed and Modular Compost, Scotts Company, UK). All pots were kept under controlled conditions in the growth chamber, bottom watered as needed during the entire experimental period and fertilized every 3 weeks (Vitax Multipurpose Soluble Feed).

Phenotypic measurements and sample collections were either taken on the same day post germination (for stomatal length, stomatal density, rosette diameter, leaf number) or at the same developmental stage (for flowering time, leaf area, trichome number, trichome density, trichome branching fruit number).

2.2.5.1 Stomatal length and stomatal density

Germination day was defined as the day when the cotyledons were first open. Seven days post germination, only two healthy seedlings were retained per pot. One of the two seedlings per pot (a total of ten seedlings per line) was retained for measuring the stomatal traits as the process was destructive and involved clipping of the cotyledons.

To obtain stomatal imprints, the cotyledons of each surplus seedling was clipped, placed on a glass slide, and the abaxial surface was covered with a thin layer of nail varnish and allowed to dry. The nail varnish was then peeled gently with the help of a clear piece of Sellotape and stuck onto a fresh labelled glass slide.

The central region of each cotyledon imprint was imaged at 400x magnification using a Nikon Confocal Microscope coupled to a Digital Sight DS- U1 colour camera (Nikon) using NIS Elements-F software. The stomatal diameter of five random stomata were measured per cotyledon using the software imageJ (Schneider *et al.*, 2012) calibrated using a stage micrometer as reference scale (5 stomatal measurements x 10 replicates x 48 lines). The average of these five measurements was used to estimate stomatal diameter per cotyledon per replicate plant. In cases where the stomatal imprint was unclear due to uneven spreading of the nail varnish layer at a microscopic level, the imprint of the second cotyledon from the pair collected was imaged. The total number of stomata in the field of view (0.212mm^2) were counted to estimate stomatal density (stomata per mm^2).

2.2.5.2 Rosette diameter and leaf number at day 15

The number of leaves were counted, and the rosette diameter of all plants were measured 15 days after germination. Rosette diameter was determined by measuring the diameter across the major axis (widest part of the rosette) and minor axis (axis perpendicular to the major axis) and averaging these values.

2.2.5.3 Flowering time

Plants were monitored daily, and flowering time was recorded as the number of days between the germination day and the appearance of first white petals.

2.2.5.4 Leaf area

To determine leaf area, the fifth rosette leaf was carefully clipped after the plants had flowered and preserved overnight in a petri-dish containing absolute ethanol. The decolourised leaf was then transferred to water in order to facilitate softening and

flattening of curled leaf edges in order to get a flat 2-dimensional leaf surface to image, thus providing an accurate image of the leaf for measurement of leaf area. A reference scale was added when each leaf was imaged. Leaf surface area was calculated using ImageJ (Schneider *et al.*, 2012).

2.2.5.5 Trichome traits

After imaging each leaf for leaf area, their adaxial surface of the leaf was observed under a dissecting microscope to count the total number of trichomes on the entire leaf. The number of branches in each trichome were counted for all trichomes on the entire leaf to calculate the percentage of 2-branched, 3-branched, 4-branched and ≥ 5-branched trichomes present on each leaf. To aid visualization of the translucent trichomes, a white LED light source was placed perpendicular to the leaf surface to allow the trichomes to reflect the light, thus appearing clearly visible and making the analysis of trichome number and branching patterns more accurate. Using the estimated leaf area from the images (described above in section 2.2.5.4), the trichome density (trichomes per cm^2) of the fifth leaf was calculated.

2.2.5.6 Fitness

At the end of the plant's life cycle (i.e. when the inflorescences stopped producing flowers and the siliques had matured), the total number of fruits on the primary and all secondary stems were counted and used as a proxy to estimate fitness.

2.2.6 Data Analysis

Statistical analysis was done using the software R version 3.5.1 (R Core Team, 2018), and all plots were generated using the ggplot2 R package (Wichkam 2016). Each phenotypic trait was analysed independently.

2.2.6.1 Determining whether colchicine has effects independent of ploidy

To determine whether colchicine affected phenotype independent of ploidy, we compared untreated diploid plants ($2x_{\text{-untreated}}$) to colchicine treated diploids ($2x_{\text{-colchicine}}$) using the following nested fixed effects ANOVA model: *Trait ~ Genotype + Treatment + Genotype:Treatment + Genotype:(Treatment/Line)*. "Treatment" indicatesd whether the plants were treated with colchicine or not; "Genotype:Treatment" was the genotype-by-treatment interaction term; and *Genotype:(Treatment/Line)* was the genotype-by-line within treatment term that estimated the variation within treated lines of the same genotype. A nested model was used wherein line was nested within

treatment because there was only one $2x_{\text{-untreated}}$ line but three $2x_{\text{-colchicine}}$ lines for each genotype. Thus, accounting for within-line variation separately in the model provided a better measure of the genotype-by-treatment effect despite the unequal sample sizes. A Tukey's HSD post-hoc test was conducted on significant genotype-by-treatment ANOVA comparisons to specifically identify genotypes in which the effects of colchicine treatment were statistically significant.

2.2.6.2 Determining whether ploidy affects phenotype and if the effects are genotype specific

Since colchicine treatment was shown to have an effect on most of the traits, it was determined that the best way to test for the effects of ploidy and genotype was to compare diploid and tetraploid lines that had both been through colchicine treatment. Thus, for all further analysis, the data from $2x_{\text{-untreated}}$ lines was discarded.

A two-way ANOVA was conducted to compare the three diploid $2x_{\text{-colchicine}}$ lines (3x10 = 30 replicates per genotype) to the three tetraploid $4x_{\text{-colchicine}}$ lines (3x10 = 30 replicates per genotype)[3] using the following nested fixed effects linear model: *Trait ~ Genotype + Ploidy + Genotype:Ploidy + (Genotype:Ploidy)/Line.* In this model, "Genotype:Ploidy" was the genotype-by-ploidy interaction term and "(Genotype:Ploidy)/Line" was line nested within genotype-by-ploidy interaction, which estimated the variation within lines of the same genotype that were at the same ploidy level (elaborated further in section 2.2.6.3). We used a nested model for the line term as each line was unique to a particular genotype and ploidy level. A Tukey's HSD post-hoc test was conducted on significant ANOVA comparisons and corrected for multiple comparisons to specifically identify the genotypes in which effects of ploidy were statistically significant.

2.2.6.3 Determining whether colchicine has any stochastic effects between independently derived lines

For all genotypes, we had multiple diploid and multiple tetraploid lines. Stochasticity within lines of the same genotype and at the same ploidy level was determined by analysing p-values obtained for the "(Genotype:Ploidy)/Line" term in the two-way ANOVA model outlined in section 2.2.6.2 above. If colchicine had no stochastic effects, the line term nested within genotype-by-ploidy would be

[3] Numbers vary for genotypes Ct, No and Wu as outlined in Table 2.2

insignificant. Obtaining significant p-values would indicate that colchicine had stochastic effects on the phenotypes.

A Tukey's HSD post-hoc test was conducted on significant ANOVA comparisons and corrected for multiple comparisons to specifically identify lines within each genotype and ploidy that were significantly different from others.

2.3 RESULTS

2.3.1 Generation and ploidy determination of colchicine treated lines

All plants treated with colchicine appeared to show arrested growth immediately after colchicine treatment, followed by abnormal rosette growth in the proceeding weeks. A large proportion of plants were weak and susceptible to fungal attack, especially powdery mildew. Overall, about 70% of the plants did not survive the treatment. Depending on the genotype, some plants either turned necrotic and died early during development, and some never transitioned to flowering and eventually decayed. Plants that survived and grew to maturity often looked stunted with some exhibiting larger flowers and increased trichome branching on the leaves.

The ploidy of G_1 plants was determined using flow cytometry. The differences in DNA content between different ploidies were indicated by their relative fluorescence intensities depicted in Figure 2.2 (A-D). Multiple sample peaks were observed either due to endoreduplication (Sugimoto-Shirasu *et al.*, 2002) in *A. thaliana* leaves, or may have corresponded to actively dividing cells in the G2 phase of the cell cycle wherein they contain twice the amount of DNA than the G1[4] phase.

[4] Note : "G1" and "G2" here refer to the G1 and G2 phases of the cell cycle respectively. These should not be confused with "G_1" and "G_2" elsewhere in the text that correspond to the first and second generation of plants post colchicine treatment.

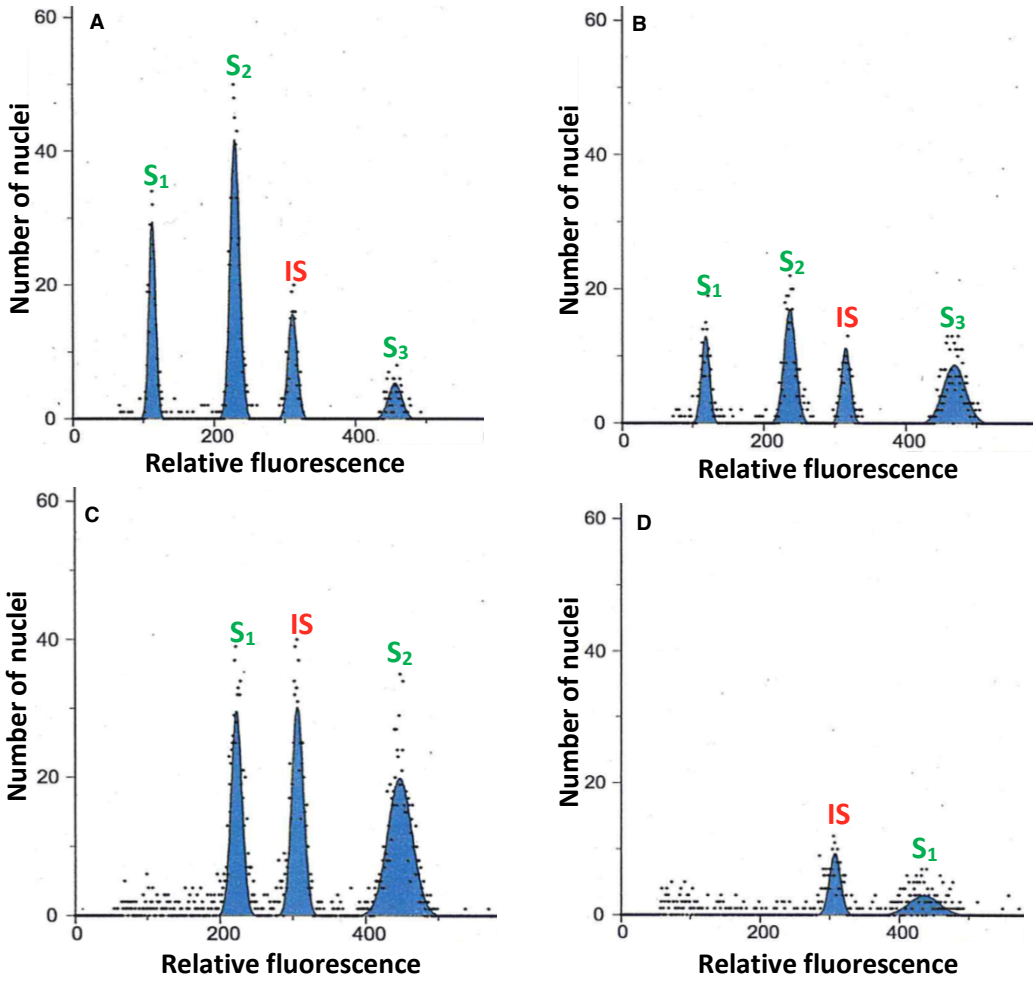


Figure 2.2 | Flow cytometric ploidy analysis of Arabidopsis thaliana leaves. Typical flow cytometry histograms of number of nuclei v/s the relative fluorescence to illustrate how sample ploidy is estimated relative to the internal standard (marked 'IS'). S1, S2 and S3 represent the first, second and third sample peaks respectively. The mean fluorescence of the first sample peak (S1) corresponds to the relative 2C DNA content of the target sample that is used to calculate its DNA content in picograms (pg) relative to the mean fluorescence of the internal standard (*Oryza sativa* leaves) that has a known 2C DNA content of 1pg. (A) Untreated diploid (2C = 2x); (B) Colchicine treated diploid (2C = 2x); (C) Colchicine treated tetraploid (2C = 4x); (D) Colchicine treated octoploid (2C = 8x).

Of the colchicine treated lines tested for ploidy, 42.2% of the plants were ascertained to be diploid, 53% tetraploid, and 4.8% octoploid. Although flow cytometry cannot reliably distinguish between tetraploids and aneuploids, the presence or absence of an extra chromosome should cause a 10% change in expected diploid DNA

content (±1 chromosome out of 10 chromosomes) and a 5% change in expected tetraploid DNA content (±1 chromosome out of 20 chromosomes). None of the DNA values we obtained fell outside these ranges, hence suggesting that all lines we obtained post colchicine treatment were likely to be euploids.

2.3.2 Colchicine affects certain phenotypic traits independent of ploidy change

A two-way ANOVA on untreated diploids ($2x_{\text{-untreated}}$) and colchicine treated diploids ($2x_{\text{-colchicine}}$) testing for the effects of colchicine treatment and genotype on ten phenotypic traits for seven genotypes showed significant genotype effects for all traits. No significant main effects of treatment were seen on any phenotypic trait; however, significant genotype-by-treatment interaction effects were seen for rosette diameter, leaf number, trichome density, stomatal length, and fitness (Table 2.3; Figures 2.3-2.7). For leaf area, percentage of trichomes with more than three branches, and stomatal density, no genotype-by-treatment effects were seen but significant line effects were observed (Table 2.3; Figures 2.8-2.10). The line effects correspond to overall variation between the colchicine treated diploid lines that belong to the same genotype. Line effects are addressed in greater detail in section 2.3.4 (that looks into variation between colchicine treated lines for the same genotype and ploidy) and are hence not discussed further at this point. Treatment was not found to have any genotype-specific or line-specific effects on flowering time, or on the total number of trichomes (Table 2.3; Figures 2.11-2.12).

Post-hoc comparisons using the Tukey's HSD test on traits that yielded significant genotype-by-treatment effects in the ANOVA showed that colchicine treated Mt diploids ($Mt2x_{\text{-colchicine}}$) had significantly higher trichome density and lower fitness (total number of fruits at maturity) than untreated Mt diploids ($Mt2x_{\text{- untreated}}$); treated Ct diploids ($Ct2x_{\text{-colchicine}}$) and No diploids ($No2x_{\text{-colchicine}}$) had significantly smaller rosette diameter than untreated Ct diploids ($Ct2x_{\text{-untreated}}$) and No diploids ($No2x_{\text{-untreated}}$) respectively; and treated Kn diploids ($Kn2x_{\text{-colchicine}}$) had significantly longer stomatal length than untreated Kn diploids ($Kn2x_{\text{-untreated}}$). Significant Tukey's HSD outputs are presented in Table 2.4. Although the ANOVA data suggested that five of the ten phenotypic traits under study had significant genotype-by-treatment effects, the post-hoc data showed that only one or two genotypes of the seven showed significant changes in response to treatment for the five phenotypes. Furthermore, the genotypes

that elicited treatment effects for one trait did not appear to show any changes in response to treatment for another trait (with the exception of Mt which showed a treatment response for two traits). This suggests that the genotypes themselves are neither susceptible nor tolerant to colchicine treatment. The combination of the trait as well as the genotype under study together determine the phenotypic changes in response to colchicine treatment.

Since treatment appears to have genotype-specific effects on phenotype even three generations after treatment, it suggests that the colchicine may have effects independent of ploidy that are heritable. Thus, when testing for the effects of ploidy in colchicine treated tetraploid lines ($4x_{\text{-colchicine}}$), colchicine treated diploid lines ($2x_{\text{-colchicine}}$) serve as a better control as any effects of colchicine will be consistent across both groups.

Table 2.3 | ANOVA data testing for the effects of colchicine between treated and untreated diploids. Two-way ANOVA results for the effect of genotype, colchicine treatment, genotype-by-treatment interaction, and line within genotype-by-treatment on various phenotypic traits. Each trait was tested independently using the fixed effects linear model *Trait ~ Genotype + Treatment + Genotype:Treatment + Genotype:(Treatment/Line)* ; N=140.

Trait	Factor	Df	Sum Sq	F value	P-value
Flowering Time (Days)	Treatment	1	32.76	3.28	7.13E-02
	Genotype	6	3877.99	64.76	**1.05E-46**
	Genotype:Treatment	6	127.60	2.13	5.08E-02
	Genotype:(Treatment/Line)	12	118.52	0.99	4.60E-01
Rosette Diameter (cm)	Treatment	1	0.63	1.23	2.69E-01
	Genotype	6	25.29	8.18	**5.10E-08**
	Genotype:Treatment	6	14.43	4.67	**1.69E-04**
	Genotype:(Treatment/Line)	12	6.93	1.12	3.43E-01
Number of Leaves (15 days post germination)	Treatment	1	0.00	0.00	9.99E-01
	Genotype	6	305.97	26.22	**1.33E-23**
	Genotype:Treatment	6	45.54	3.90	**9.91E-04**
	Genotype:(Treatment/Line)	12	24.59	1.05	4.01E-01
Leaf Area (cm2)	Treatment	1	0.66	1.63	2.03E-01
	Genotype	6	13.43	5.56	**2.16E-05**
	Genotype:Treatment	6	3.88	1.61	1.46E-01

	Genotype:(Treatment/Line)	12	9.28	1.92	**3.32E-02**
Trichome Density (trichomes per cm2)	Treatment	1	166.27	0.84	3.59E-01
	Genotype	6	88074.10	74.56	**2.21E-50**
	Genotype:Treatment	6	3957.03	3.35	**3.53E-03**
	Genotype:(Treatment/Line)	12	3577.96	1.51	1.20E-01
Total Trichomes (per leaf)	Treatment	1	10.82	0.01	9.28E-01
	Genotype	6	411104.05	51.35	**2.07E-39**
	Genotype:Treatment	6	11460.57	1.43	2.04E-01
	Genotype:(Treatment/Line)	12	8957.04	0.56	8.73E-01
Trichomes with >3 branches (% of total trichomes on the leaf)	Treatment	1	0.21	0.00	9.45E-01
	Genotype	6	27621.83	106.68	**1.89E-62**
	Genotype:Treatment	6	361.56	1.40	2.17E-01
	Genotype:(Treatment/Line)	12	1017.59	1.96	**2.85E-02**
Stomatal Density (per mm2)	Treatment	1	1508.67	1.27	2.60E-01
	Genotype	6	124434.30	17.50	**2.49E-16**
	Genotype:Treatment	6	6861.49	0.97	4.50E-01
	Genotype:(Treatment/Line)	12	49019.02	3.45	**1.23E-04**
Stomatal Diameter (µm)	Treatment	1	3.08	0.51	4.75E-01
	Genotype	6	99.82	2.77	**1.32E-02**
	Genotype:Treatment	6	125.10	3.47	**2.77E-03**
	Genotype:(Treatment/Line)	12	106.40	1.47	1.36E-01
Total Fruits	Treatment	1	48456.06	6.33	1.26E-02
	Genotype	6	501737.14	10.92	**1.07E-10**
	Genotype:Treatment	6	90392.69	1.97	**7.12E-02**
	Genotype:(Treatment/Line)	12	105782.94	1.15	3.20E-01

Table 2.4 | Output table of Tukey's HSD test for ANOVAs that tested significant for the effects of genotype-by-treatment (Genotype:Treatment) for ten phenotypic traits. 91 Tukey comparisons were obtained for each trait for all combinations of genotypes and treatments. Only the significant Tukey comparisons within the same genotype have been presented here. Lower and upper bound represent the lower and upper limits of the 95% confidence interval. Sig represents the adjusted P-value for the comparison corrected for multiple comparisons.

Trait	$2x_{\text{-colchicine}}$ (I)	$2x_{\text{-untreated}}$ (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Rosette Diameter (cm)	Ct $2x_{\text{-colchicine}}$	Ct $2x_{\text{-untreated}}$	-0.97	-1.92	-0.02	0.042
Rosette Diameter (cm)	No $2x_{\text{-colchicine}}$	No $2x_{\text{-untreated}}$	-0.97	-1.93	-0.01	0.046

Trait	$2x_{\text{-colchicine}}$ (I)	$2x_{\text{-untreated}}$ (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Trichome Density (trichomes per cm^2)	Mt $2x_{\text{-colchicine}}$	Mt $2x_{\text{-untreated}}$	19.81	2.43	37.19	0.010
Stomatal Diameter (µm)	Kn $2x_{\text{-colchicine}}$	Kn $2x_{\text{-untreated}}$	5.66	0.59	10.73	0.014
Total Fruits	Mt $2x_{\text{-colchicine}}$	Mt $2x_{\text{-untreated}}$	-128.43	-236.75	-20.12	0.006

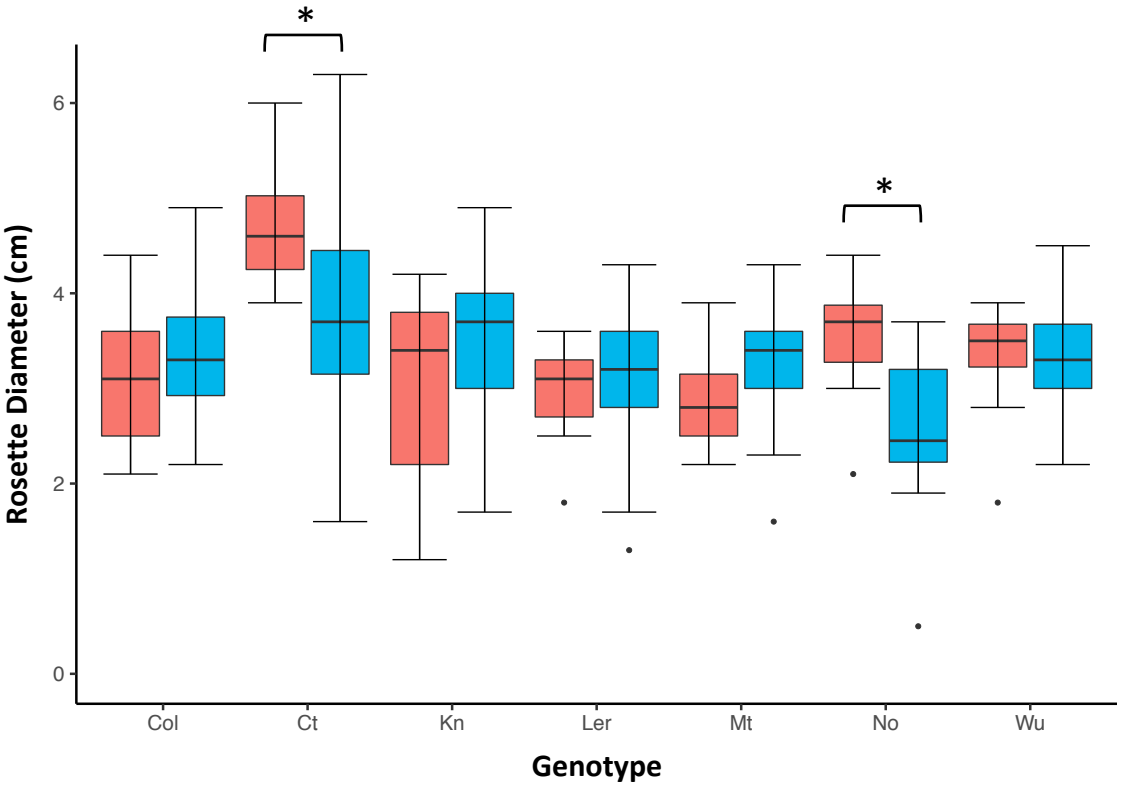


Figure 2.3 | Rosette diameter data for untreated and colchicine treated diploids. Rosette diameter was measured along two perpendicular rosette axes and averaged. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median rosette diameter, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers. Significance (*) : adjusted p-value<0.05 using a Tukey's HSD test corrected for multiple comparisons.

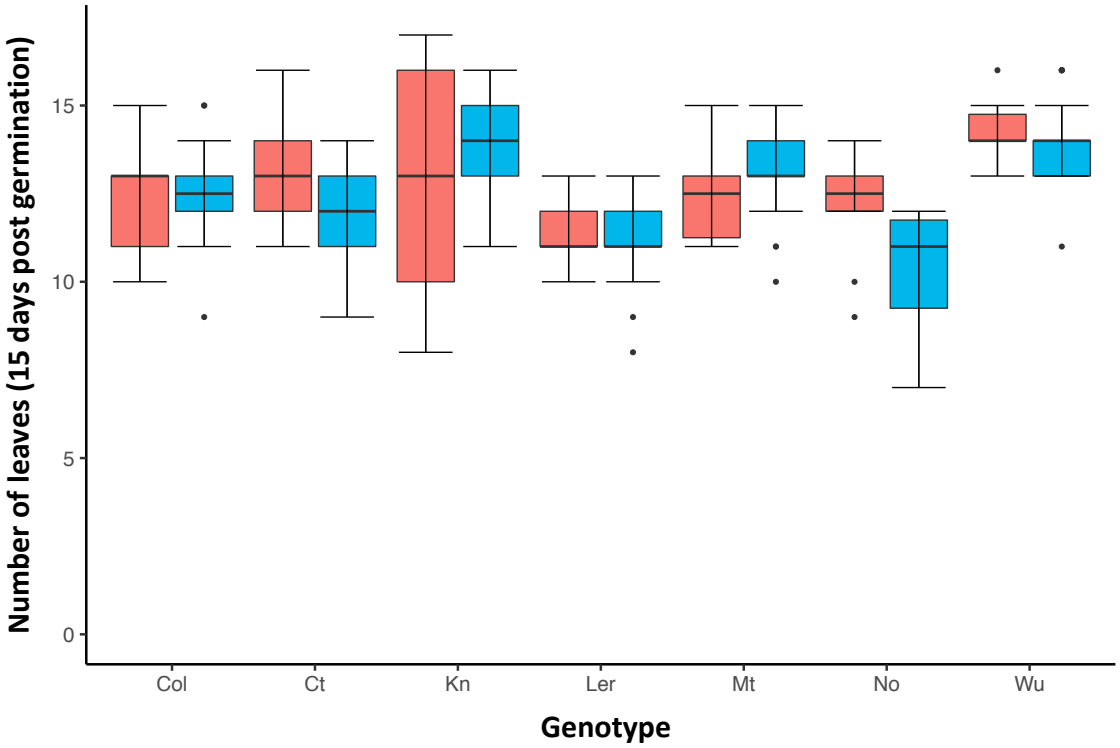


Figure 2.4 | Leaf number data for untreated and colchicine treated diploids. The number of rosette leaves were counted 15 days post germination. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median leaf number, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers.

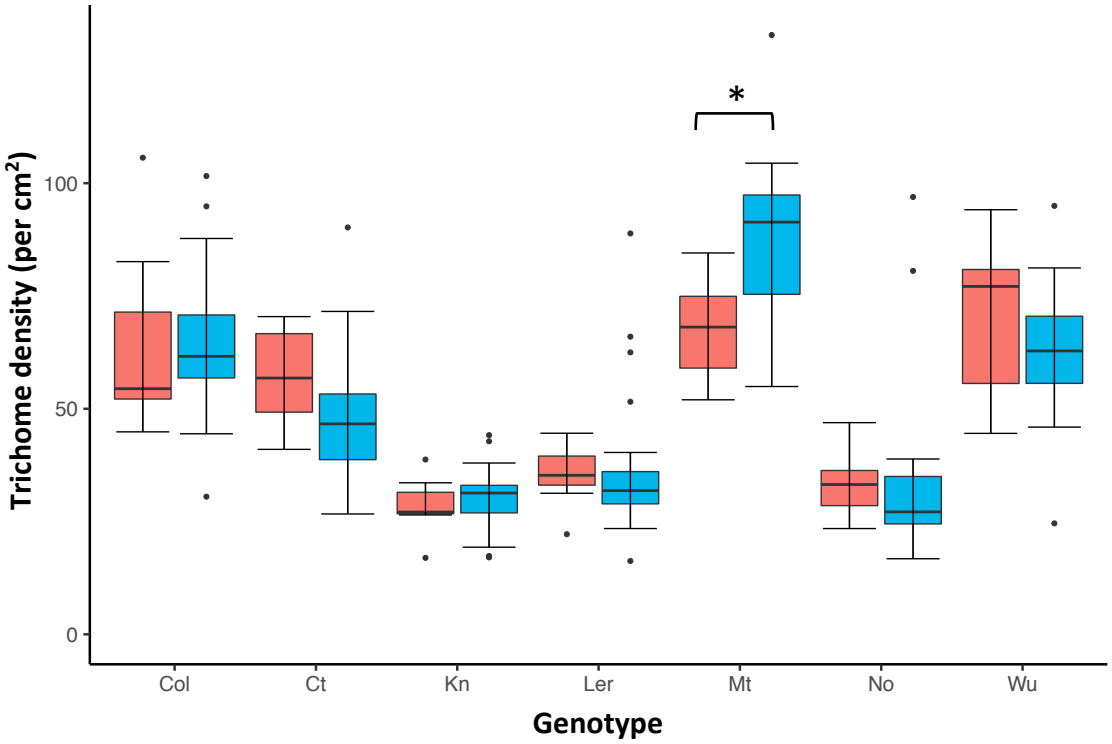


Figure 2.5 | Trichome density data for untreated and colchicine treated diploids. Trichome density for the 5th rosette leaf was calculated as the number of trichomes per cm^2 of the leaf. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median trichome density, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers. Significance (*) : adjusted p-value<0.05 using a Tukey's HSD test corrected for multiple comparisons.

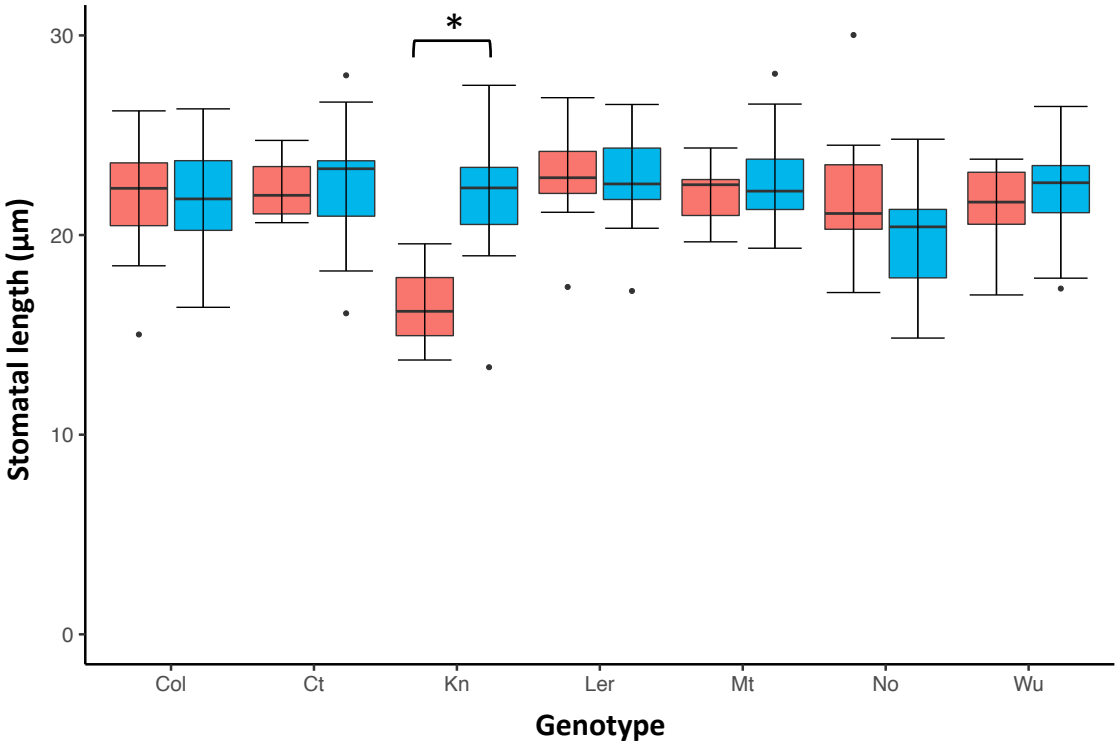


Figure 2.6 | Stomatal length data for untreated and colchicine treated diploids. Stomatal length of stomata in the central region of the cotyledons of two-week old seedlings was measured as the average distance between the ends of five open stoma for each plant. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median stomatal length, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers. Significance (*) : adjusted p-value<0.05 using a Tukey's HSD test corrected for multiple comparisons.

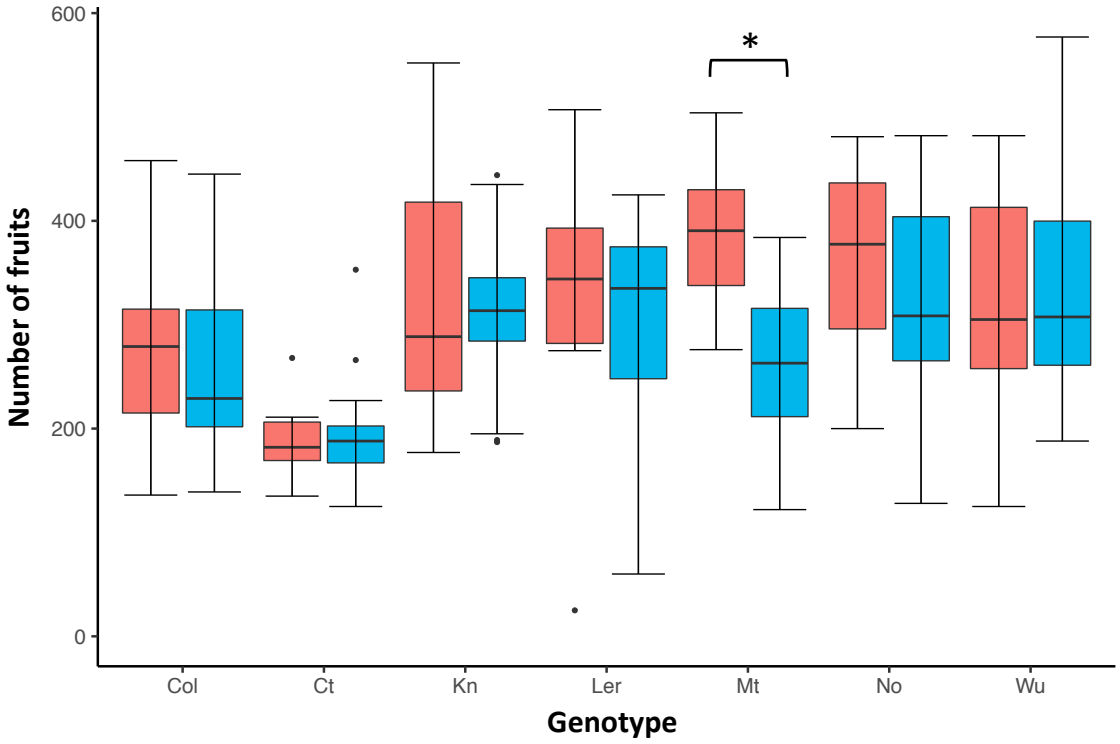


Figure 2.7 | Fruit number data for untreated and colchicine treated diploids. The total number of fruits on each plant were counted at the end of the plant's life cycle. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median fruit number, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers. Significance (*) : adjusted p-value<0.05 using a Tukey's HSD test corrected for multiple comparisons.

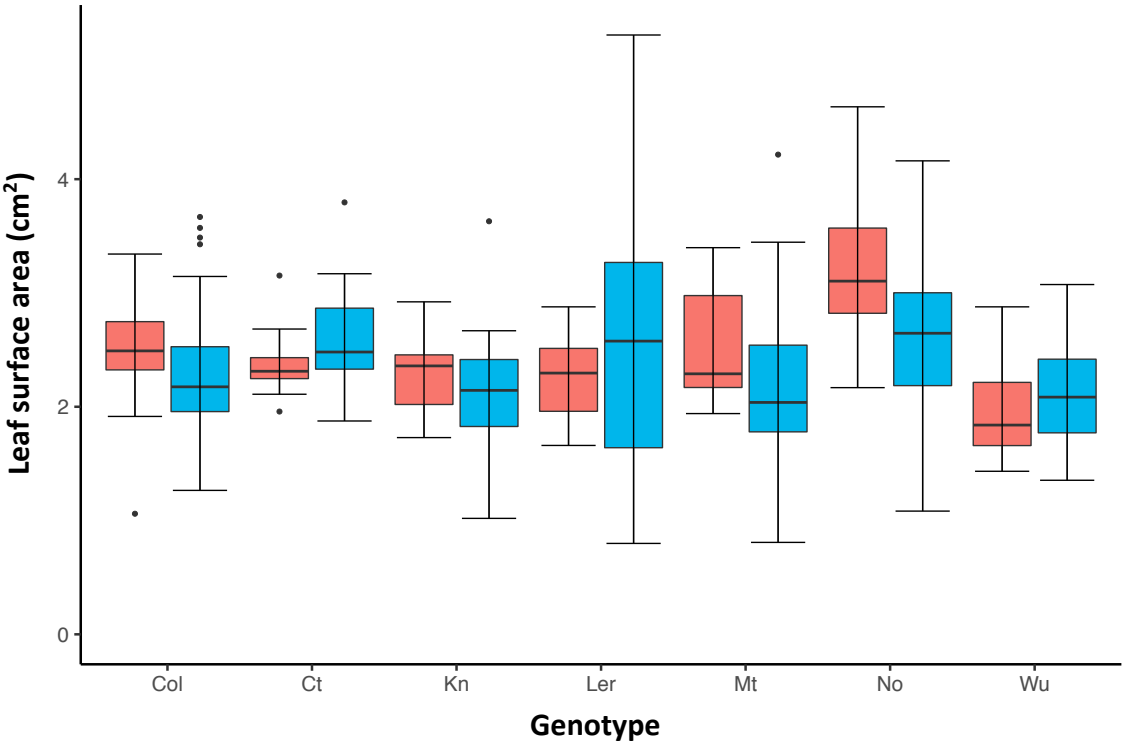


Figure 2.8 | Leaf surface area data for untreated and colchicine treated diploids. Leaf area of the 5th rosette leaf was measured post flowering. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median leaf area, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers.

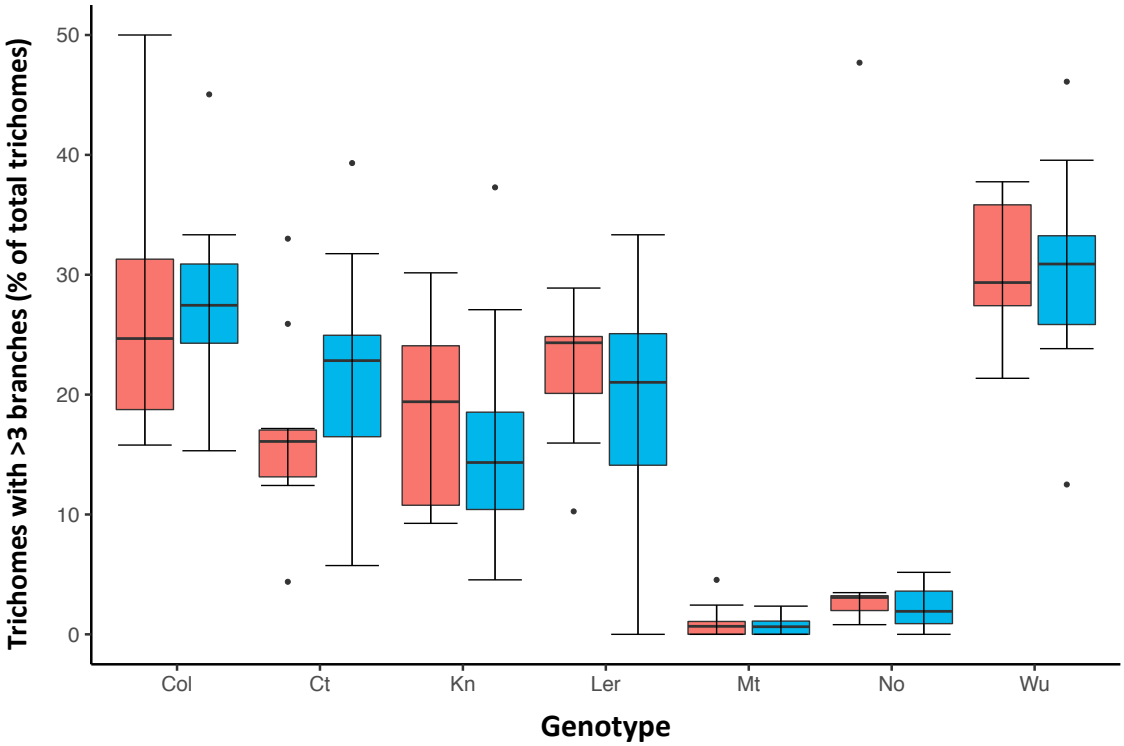


Figure 2.9 | Trichome branching (>3 branches) data for untreated and colchicine treated diploids. The number of branches on each trichome on adaxial surface of the 5th rosette leaf were counted post flowering and the percentage of trichomes that had more than three branches (of the total number of trichomes) was calculated. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median percentage of trichomes having more than three branches, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers.

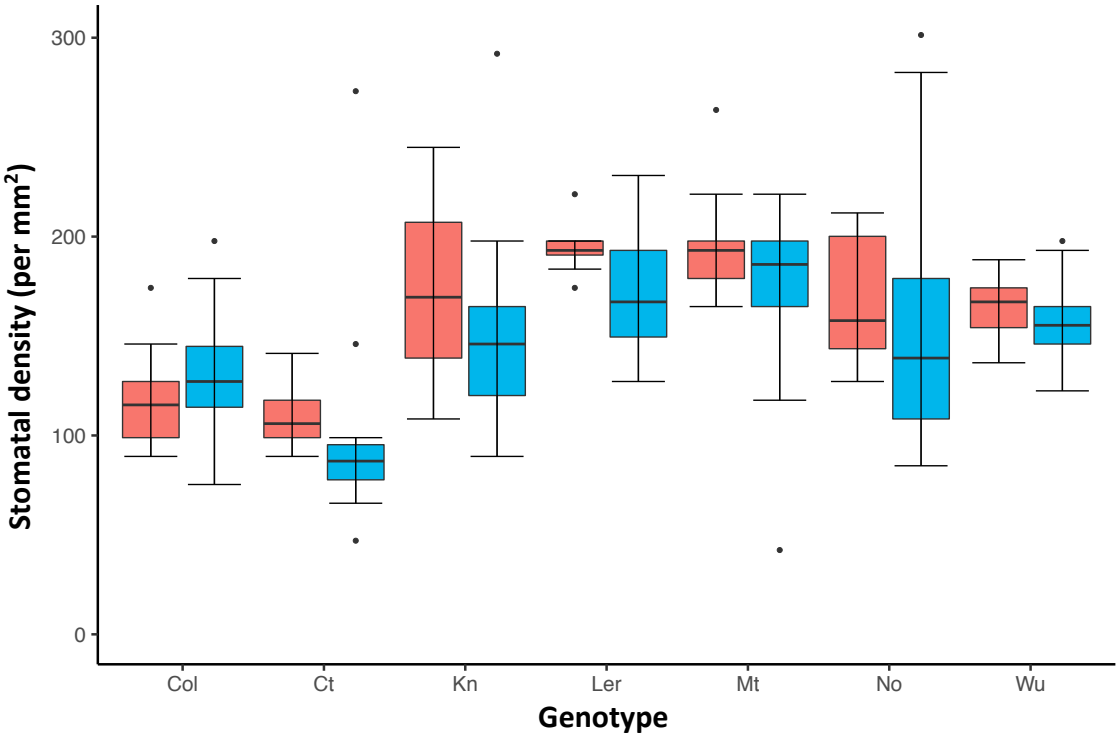


Figure 2.10 | Stomatal density data for untreated and colchicine treated diploids. Stomatal density in the central region of the cotyledons of two-week old seedlings was calculated as the number of stomata per mm^2 of the cotyledon. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median stomatal density, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers.

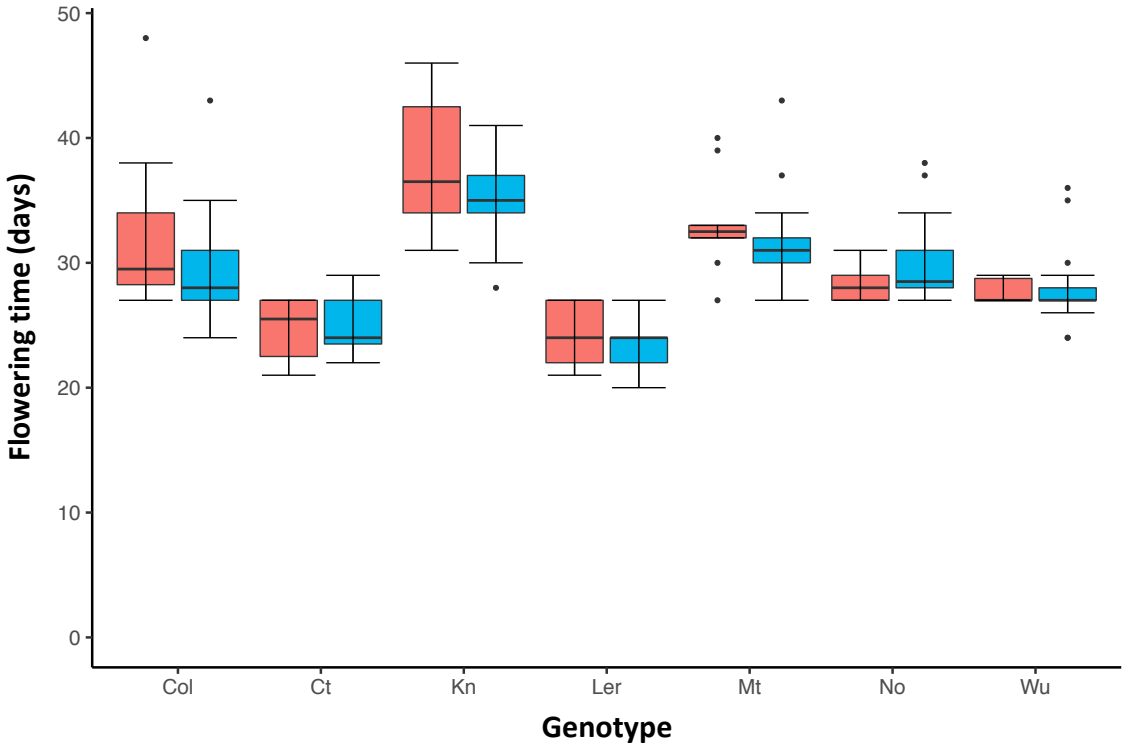


Figure 2.11 | Flowering time and rosette diameter data for untreated and colchicine treated diploids. Flowering time was measured in days taken for the appearance of the first white petals post germination for seven genotypes. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median flowering time, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers.

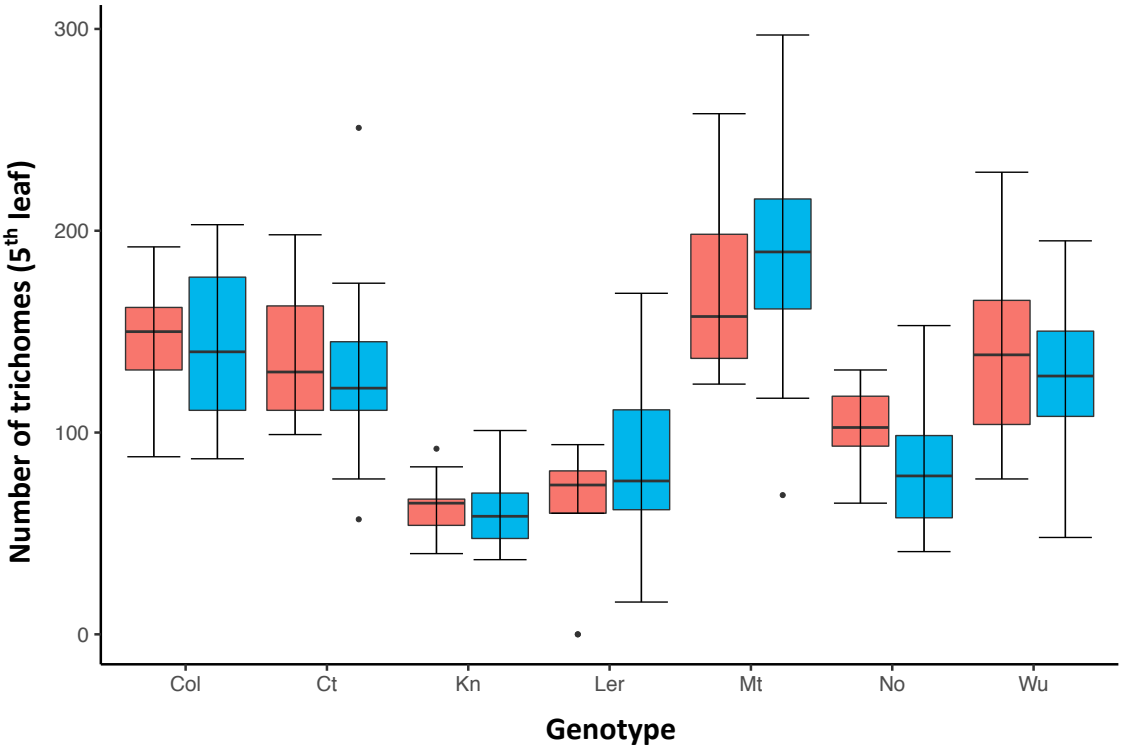


Figure 2.12 | Trichome number data for untreated and colchicine treated diploids. The total number of trichomes on adaxial surface of the 5th rosette leaf were counted post flowering. Red boxes represent untreated diploids (N=10 for each genotype) and blue boxes represent colchicine treated diploids (N=30 for each genotype). Horizontal lines within the boxes represent the median trichome number, box extremities represent the interquartile range, whiskers indicate the minimum and maximum values and black dots represent potential outliers.

2.3.3 Phenotypic changes are seen in response to increased ploidy

Diploid and tetraploid lines that had been exposed to colchicine ($2x_{\text{-colchicine}}$ and $4x_{\text{-colcicine}}$) were used to analyse ploidy effects. Summary statistics are available in Supplementary Data 7.2. A two-way ANOVA on the phenotypic data analysed for each trait independently showed that significant genotype, ploidy and genotype-by-ploidy interaction effects were observed for all traits (Table 2.5). A Tukey's HSD post hoc test was conducted to further investigate specific genotypes that showed differences between diploids and tetraploids (Table 2.6), and to determine if the direction and magnitude of change in phenotypic response to ploidy were consistent or varied across genotypes.

Flowering time data revealed that of the seven genotypes, only Kn plants showed

significant differences between diploids and tetraploids. Kn tetraploids flowered an average of 2.9 days later than Kn diploids (Figure 2.13; Table 2.6). Rosette diameter data revealed that tetraploids for genotypes Ct, No and Wu had significantly longer rosette diameters 15 days post germination than their diploid controls (Figure 2.14; Table 2.6).

Kn, Ler and Mt tetraploids had significantly lesser rosette leaves 15 days post germination than their respective diploids (Figure 2.15; Table 2.6).

The 5th rosette leaves of tetraploid Kn, No and Wu collected at the time of flowering had significantly larger leaf surface areas than their corresponding diploids (Figure 2.16; Table 2.6).

Tetraploids of all seven genotypes showed a significant increase in trichome branching (i.e. increase in the percentage of trichomes that had more than three branches) (Figure 2.17; Table 2.6). In terms of the number of trichomes on the leaf, only Wu tetraploids had significantly more trichomes in comparison to Wu diploids (Figure 2.18; Table 2.6). The trichome density in tetraploids was significantly higher, overall, in comparison to diploids (Supplementary Table S7.2), however, no specific genotype showed any significant difference in trichome density (Figure 2.19; Table 2.6)

Tetraploids of all seven genotypes showed significantly decreased stomatal density and significantly increased stomatal length when compared to their respective diploid genotypes (Figure 2.20-2.21; Table 2.6).

Col and No tetraploids showed significantly lower fitness than diploids (measured as the total number of fruits at the end of the plant's life cycle) (Figure 2.22; Table 2.6).

Overall, we found that the response to increased ploidy was genotype dependent. For traits like flowering time, total trichome number, and fruit number, only one or two of the seven genotypes exhibited significant phenotypic differences. For other traits like trichome branching, stomatal density, and stomatal length, all genotypes showed significant differences between diploids and tetraploids. Furthermore, no single genotype showed a significant ploidy-response to all traits, and the genotypes that showed significant ploidy-related changes for one trait oftentimes did not overlap with the genotypes that showed significant responses for another trait (e.g. Col and Kn showed phenotypic changes in fitness in response to ploidy, but did not show any change in leaf number, where the effects in Kn, Ler, and Mt were more prominent).

2.3.4 Colchicine treatment has stochastic effects on some phenotypes

To assess if the effects of colchicine were stochastic, we tested for differences in phenotypes between the multiple independently derived lines at each ploidy level (diploid and tetraploid) within each genotype using a nested fixed effects linear model ANOVA. The ANOVA data showed significant differences between lines of the same genotype and ploidy level for leaf area, trichome density, percentage of trichomes with more than 3 branches, stomatal density and fruit number (Table 2.3; (Genotype x Ploidy)/Line term). However, the Tukey's HSD test conducted for phenotypes that showed significant (Genotype x Ploidy)/Line term in their ANOVA output showed that significant differences were observed for only one genotype (between line Col 4x – B and Col 4x – C) for one trait - percentage of trichomes with more than 3 branches (Table 2.5; Supplementary Data 7.3; Figure 2.17). As the statistical significance appeared to be lost for most traits after correcting for multiple comparisons with the Tukey's HSD test (Table S2.4), our data suggested that, overall, colchicine did not have any stochastic effects on the phenotypes studied i.e. all the lines we obtained from independent colchicine treatment events for the same genotype and at the same ploidy level are essentially phenotypically equivalent (with the only exception being Col 4x – B and Col 4x – C for trichome branching).

Table 2.5 | ANOVA results for the effect of genotype and colchicine on various phenotypic traits for diploid untreated and colchicine treated plants for seven natural *A. thaliana* genotypes. Two-way ANOVA results for the effect of genotype, ploidy, genotype-by-ploidy interaction, and line within genotype-by-ploidy on various phenotypic traits. Each trait was tested independently using the fixed effects linear model *Trait ~ Genotype + Ploidy + Genotype:Ploidy + (Genotype:Ploidy)/Line)* ; N=410.

Trait	Factor	Df	Sum Sq	F value	P-value
Flowering Time (Days)	Genotype	6	5423.54	123.88	**2.64E-84**
	Ploidy	1	32.65	4.48	**3.51E-02**
	Genotype:Ploidy	6	235.74	5.38	**2.42E-05**
	(Genotype:Ploidy)/Line	27	198.10	1.01	4.60E-01
Rosette Diameter (cm)	Genotype	6	71.77	28.50	**7.90E-28**
	Ploidy	1	15.69	37.38	**2.55E-09**
	Genotype:Ploidy	6	20.04	7.96	**4.55E-08**
	(Genotype:Ploidy)/Line	27	10.89	0.96	5.23E-01

Trait	Factor	Df	Sum Sq	F value	P-value
Number of Leaves (15 days post germination)	Genotype	6	608.43	66.95	**1.66E-55**
	Ploidy	1	35.81	23.64	**1.74E-06**
	Genotype:Ploidy	6	68.73	7.56	**1.19E-07**
	(Genotype:Ploidy)/Line	27	42.01	1.03	4.30E-01
Leaf Area (cm2)	Genotype	6	29.64	10.01	**3.50E-10**
	Ploidy	1	20.37	41.27	**4.48E-10**
	Genotype:Ploidy	6	12.14	4.10	**5.49E-04**
	(Genotype:Ploidy)/Line	27	25.21	1.89	**5.51E-03**
Trichome Density (trichomes per cm2)	Genotype	6	158576.37	106.40	**2.36E-74**
	Ploidy	1	1300.80	5.24	**2.27E-02**
	Genotype:Ploidy	6	6620.27	4.44	**2.43E-04**
	(Genotype:Ploidy)/Line	27	10624.18	1.58	**3.50E-02**
Total Trichomes (per leaf)	Genotype	6	711760.95	79.55	**2.47E-61**
	Ploidy	1	10238.41	6.87	**9.19E-03**
	Genotype:Ploidy	6	101580.93	11.35	**1.50E-11**
	(Genotype:Ploidy)/Line	27	38633.02	0.96	5.26E-01
Trichomes with >3 branches (% of total trichomes)	Genotype	6	72572.62	97.95	**1.67E-70**
	Ploidy	1	172508.03	1396.99	**3.46E-121**
	Genotype:Ploidy	6	20441.41	27.59	**1.05E-26**
	(Genotype:Ploidy)/Line	27	8620.95	2.59	**4.50E-05**
Stomatal Density (per mm2)	Genotype	6	159780.19	36.29	**8.03E-34**
	Ploidy	1	445750.27	607.43	**2.55E-77**
	Genotype:Ploidy	6	15980.40	3.63	**1.67E-03**
	(Genotype:Ploidy)/Line	27	55966.18	2.82	**7.48E-06**
Stomatal Diameter (µm)	Genotype	6	109.21	2.31	**3.36E-02**
	Ploidy	1	9434.72	1198.16	**2.37E-112**
	Genotype:Ploidy	6	212.11	4.49	**2.17E-04**
	(Genotype:Ploidy)/Line	27	238.30	1.12	3.12E-01
Total Fruits	Genotype	6	861554.02	23.45	**2.13E-23**
	Ploidy	1	51586.75	8.42	**3.93E-03**
	Genotype:Ploidy	6	357929.40	9.74	**5.94E-10**
	(Genotype:Ploidy)/Line	27	330134.76	2.00	**2.69E-03**

Table 2.6 | Output table of Tukey's HSD test for ANOVAs that tested significant for the effects of genotype-by-ploidy (Genoptye:Ploidy) for ten phenotypic traits. 91 Tukey comparisons were obtained for each trait for all combinations of genotypes and ploidies. Ploidy comparisons within the same genotype have been presented here (seven comparisons). Lower and upper bound represent the lower and upper limits of the 95% confidence interval. Sig represents the adjusted p-value for the comparison corrected for multiple comparisons.

Trait	4x-colchicine (I)	2x-colchicine (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Flowering Time (Days)	Col 4x-colchicine	Col 2x-colchicine	-1.867	-4.222	0.489	0.297
	Ct 4x-colchicine	Ct 2x-colchicine	1.578	-0.964	4.119	0.704
	Kn 4x-colchicine	**Kn 2x-colchicine**	**2.894**	**0.455**	**5.333**	**0.006**
	Ler 4x-colchicine	Ler 2x-colchicine	2.036	-0.428	4.499	0.234
	Mt 4x-colchicine	Mt 2x-colchicine	0.200	-2.155	2.555	1.000
	No 4x-colchicine	No 2x-colchicine	0.200	-2.298	2.698	1.000
	Wu 4x-colchicine	Wu 2x-colchicine	-0.867	-3.500	1.767	0.998
Rosette Diameter (cm)	Col 4x-colchicine	Col 2x-colchicine	0.373	-0.192	0.938	0.606
	Ct 4x-colchicine	**Ct 2x-colchicine**	**0.855**	**0.245**	**1.464**	**0.000**
	Kn 4x-colchicine	Kn 2x-colchicin	0.461	-0.109	1.030	0.266
	Ler 4x-colchicine	Ler 2x-colchicine	-0.375	-0.966	0.215	0.669
	Mt 4x-colchicine	Mt 2x-colchicine	-0.070	-0.649	0.510	1.000
	No 4x-colchicine	**No 2x-colchicine**	**1.005**	**0.384**	**1.626**	**0.000**
	Wu 4x-colchicine	**Wu 2x-colchicine**	**0.738**	**0.107**	**1.370**	**0.007**
Number of Leaves (15 days post germination)	Col 4x-colchicine	Col 2x-colchicine	-0.800	-1.873	0.273	0.399
	Ct 4x-colchicine	Ct 2x-colchicine	0.661	-0.497	1.819	0.811
	Kn 4x-colchicine	**Kn 2x-colchicine**	**-1.376**	**-2.458**	**-0.294**	**0.002**
	Ler 4x-colchicine	**Ler 2x-colchicine**	**-1.711**	**-2.833**	**-0.588**	**0.000**
	Mt 4x-colchicine	**Mt 2x-colchicine**	**-1.172**	**-2.274**	**-0.071**	**0.025**
	No 4x-colchicine	No 2x-colchicine	0.356	-0.824	1.535	0.999
	Wu 4x-colchicine	Wu 2x-colchicine	0.100	-1.100	1.300	1.000
Leaf Area (cm2)	Col 4x-colchicine	Col 2x-colchicine	-0.043	-0.661	0.575	1.000
	Ct 4x-colchicine	Ct 2x-colchicine	0.172	-0.498	0.842	1.000
	Kn 4x-colchicine	**Kn 2x-colchicine**	**0.991**	**0.331**	**1.652**	**0.000**
	Ler 4x-colchicine	Ler 2x-colchicine	0.334	-0.326	0.995	0.911
	Mt 4x-colchicine	Mt 2x-colchicine	0.316	-0.308	0.940	0.911
	No 4x-colchicine	**No 2x-colchicine**	**0.808**	**0.143**	**1.474**	**0.004**
	Wu 4x-colchicine	**Wu 2x-colchicine**	**0.840**	**0.144**	**1.536**	**0.004**
Trichome Density (trichomes per cm2)	Col 4x-colchicine	Col 2x-colchicine	0.421	-13.563	14.406	1.000
	Ct 4x-colchicine	Ct 2x-colchicine	-2.329	-17.664	13.007	1.000
	Kn 4x-colchicine	Kn 2x-colchicine	-6.314	-21.388	8.760	0.980
	Ler 4x-colchicine	Ler 2x-colchicine	-9.029	-23.842	5.785	0.728

Trait	4x-colchicine (I)	2x-colchicine (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
	Mt 4x-colchicine	Mt 2x-colchicine	-13.558	-27.551	0.435	0.068
	No 4x-colchicine	No 2x-colchicine	-8.479	-23.406	6.447	0.815
	Wu 4x-colchicine	Wu 2x-colchicine	15.305	-0.308	30.918	0.061
Total Trichomes (per leaf)	Col 4x-colchicine	Col 2x-colchicine	1.241	-33.024	35.507	1.000
	Ct 4x-colchicine	Ct 2x-colchicine	-4.427	-42.003	33.149	1.000
	Kn 4x-colchicine	Kn 2x-colchicine	6.670	-30.265	43.605	1.000
	Ler 4x-colchicine	Ler 2x-colchicine	-9.964	-46.260	26.332	1.000
	Mt 4x-colchicine	Mt 2x-colchicine	-7.531	-41.817	26.755	1.000
	No 4x-colchicine	No 2x-colchicine	2.279	-34.296	38.853	1.000
	Wu 4x-colchicine	**Wu 2x-colchicine**	**96.726**	**58.470**	**134.983**	**0.000**
Trichomes with >3 branches (% of total trichomes on the leaf)	**Col 4x-colchicine**	**Col 2x-colchicine**	**35.005**	**25.145**	**44.866**	**0.000**
	Ct 4x-colchicine	**Ct 2x-colchicine**	**39.796**	**28.983**	**50.608**	**0.000**
	Kn 4x-colchicine	**Kn 2x-colchicine**	**68.604**	**57.976**	**79.232**	**0.000**
	Ler 4x-colchicine	**Ler 2x-colchicine**	**54.016**	**43.571**	**64.460**	**0.000**
	Mt 4x-colchicine	**Mt 2x-colchicine**	**18.392**	**8.527**	**28.258**	**0.000**
	No 4x-colchicine	**No 2x-colchicine**	**42.988**	**32.463**	**53.512**	**0.000**
	Wu 4x-colchicine	**Wu 2x-colchicine**	**51.809**	**40.801**	**62.818**	**0.000**
Stomatal Density (per mm2)	**Col 4x-colchicine**	**Col 2x-colchicine**	**-64.981**	**-88.815**	**-41.147**	**0.000**
	Ct 4x-colchicine	**Ct 2x-colchicine**	**-36.611**	**-66.734**	**-6.487**	**0.004**
	Kn 4x-colchicine	**Kn 2x-colchicine**	**-83.572**	**-107.850**	**-59.293**	**0.000**
	Ler 4x-colchicine	**Ler 2x-colchicine**	**-83.412**	**-110.616**	**-56.209**	**0.000**
	Mt 4x-colchicine	**Mt 2x-colchicine**	**-76.854**	**-101.780**	**-51.927**	**0.000**
	No 4x-colchicine	**No 2x-colchicine**	**-72.123**	**-98.099**	**-46.146**	**0.000**
	Wu 4x-colchicine	**Wu 2x-colchicine**	**-73.535**	**-99.955**	**-47.115**	**0.000**
Stomatal Diameter (µm)	**Col 4x-colchicine**	**Col 2x-colchicine**	**11.362**	**8.893**	**13.831**	**0.000**
	Ct 4x-colchicine	**Ct 2x-colchicine**	**7.992**	**4.872**	**11.113**	**0.000**
	Kn 4x-colchicine	**Kn 2x-colchicine**	**10.623**	**8.107**	**13.138**	**0.000**
	Ler 4x-colchicine	**Ler 2x-colchicine**	**9.266**	**6.448**	**12.085**	**0.000**
	Mt 4x-colchicine	**Mt 2x-colchicine**	**8.589**	**5.957**	**11.221**	**0.000**
	No 4x-colchicine	**No 2x-colchicine**	**11.565**	**8.864**	**14.267**	**0.000**
	Wu 4x-colchicine	**Wu 2x-colchicine**	**12.823**	**10.086**	**15.560**	**0.000**
Total Fruits	**Col 4x-colchicine**	**Col 2x-colchicine**	**-110.114**	**-178.929**	**-41.298**	**0.000**
	Ct 4x-colchicine	Ct 2x-colchicine	23.563	-50.368	97.493	0.998
	Kn 4x-colchicine	Kn 2x-colchicine	-65.700	-133.930	2.530	0.073
	Ler 4x-colchicine	Ler 2x-colchicine	56.444	-14.926	127.814	0.300
	Mt 4x-colchicine	Mt 2x-colchicine	42.667	-25.563	110.896	0.693
	No 4x-colchicine	**No 2x-colchicine**	**-77.575**	**-149.944**	**-5.206**	**0.023**
	Wu 4x-colchicine	Wu 2x-colchicine	-25.267	-101.550	51.017	0.998

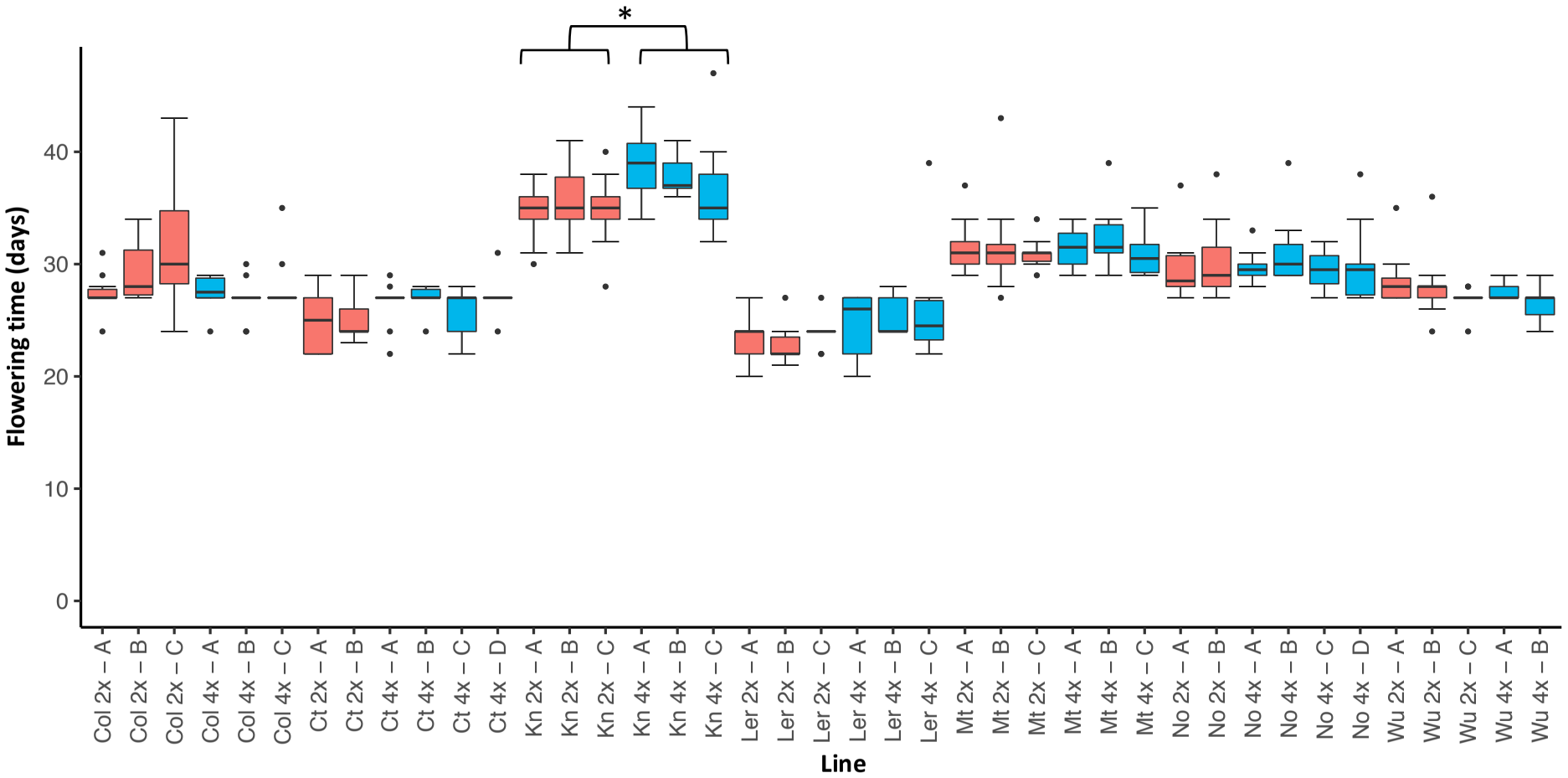


Figure 2.13 | Flowering time data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers. Significance (*) : adjusted p-value < 0.05 using a Tukey's HSD test corrected for multiple comparisons.

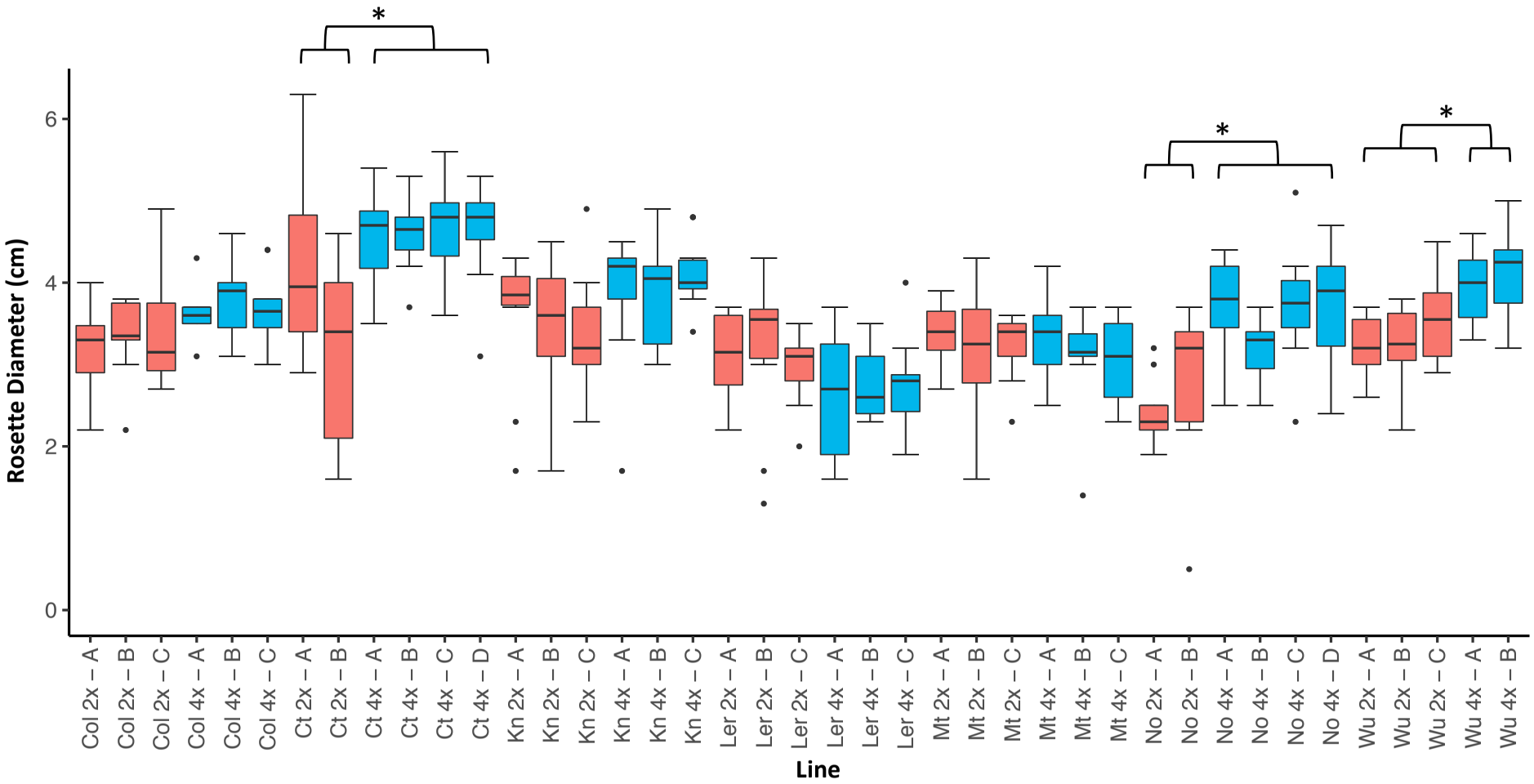


Figure 2.14 | Rosette diameter data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers. Significance (*) : adjusted p-value < 0.05 using a Tukey's HSD test corrected for multiple comparisons.

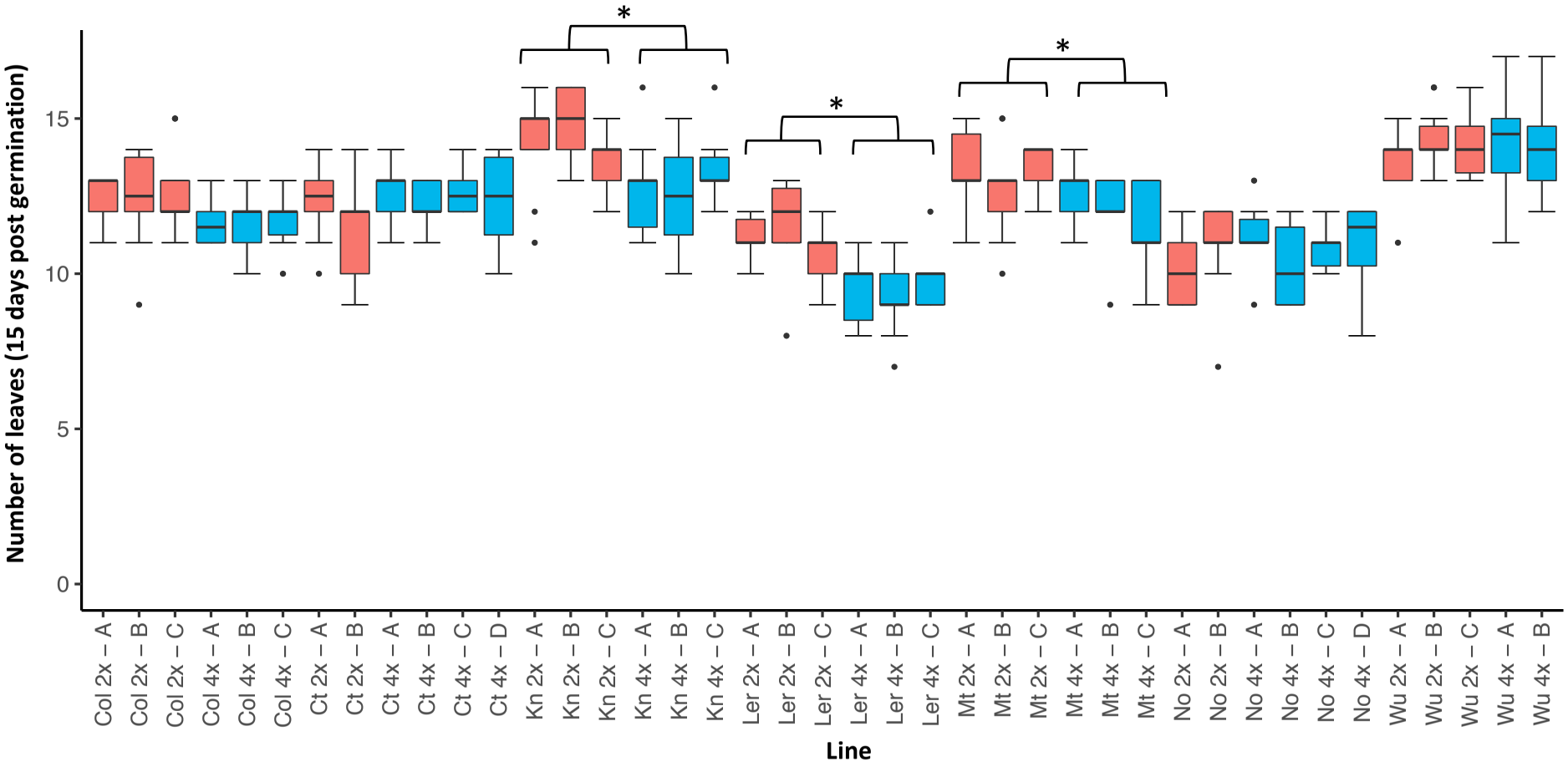


Figure 2.15 | Leaf number data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers. Significance (*) : adjusted p-value < 0.05 using a Tukey's HSD test corrected for multiple comparisons.

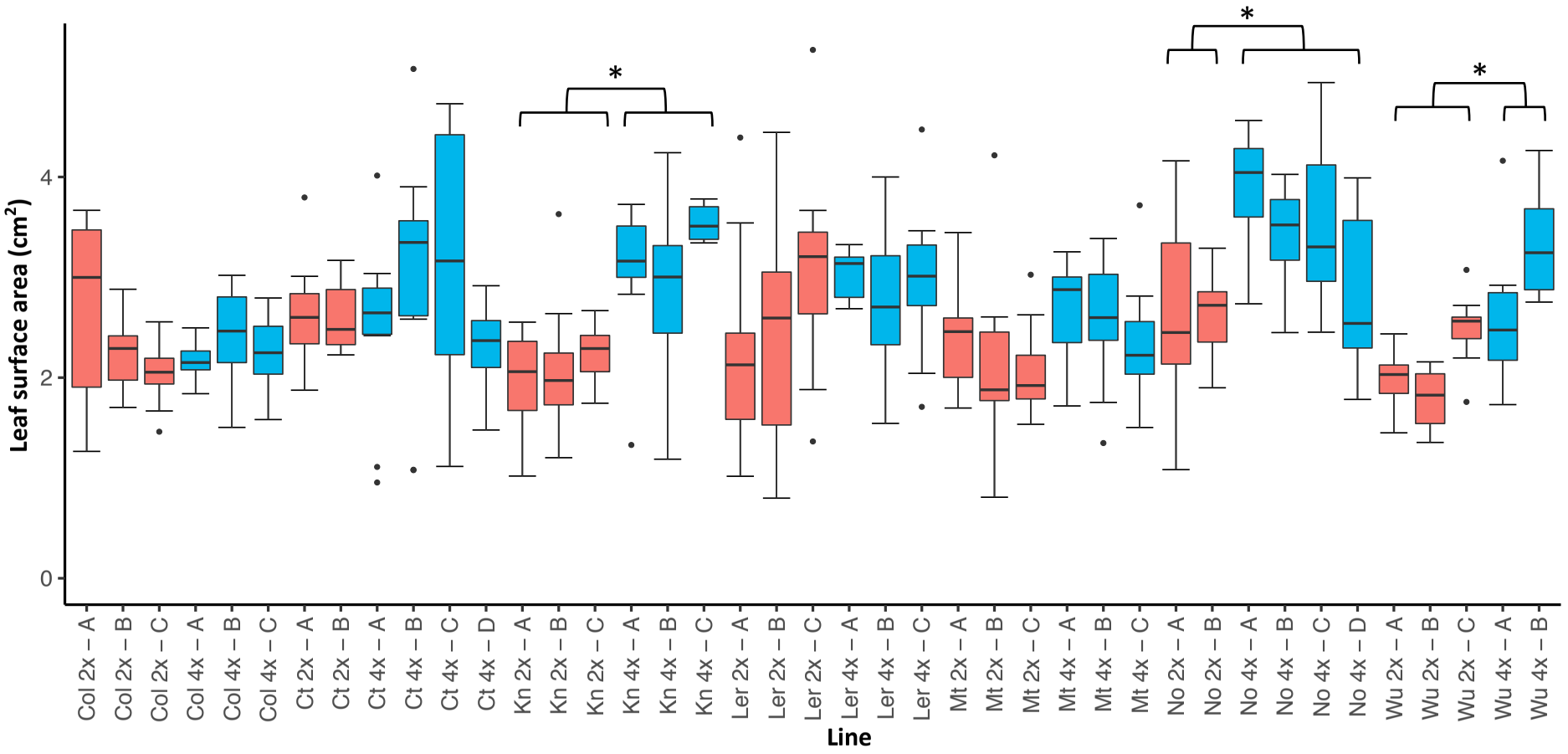


Figure 2.16 | Leaf surface area data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers. Significance (*) : adjusted p-value < 0.05 using a Tukey's HSD test corrected for multiple comparisons.

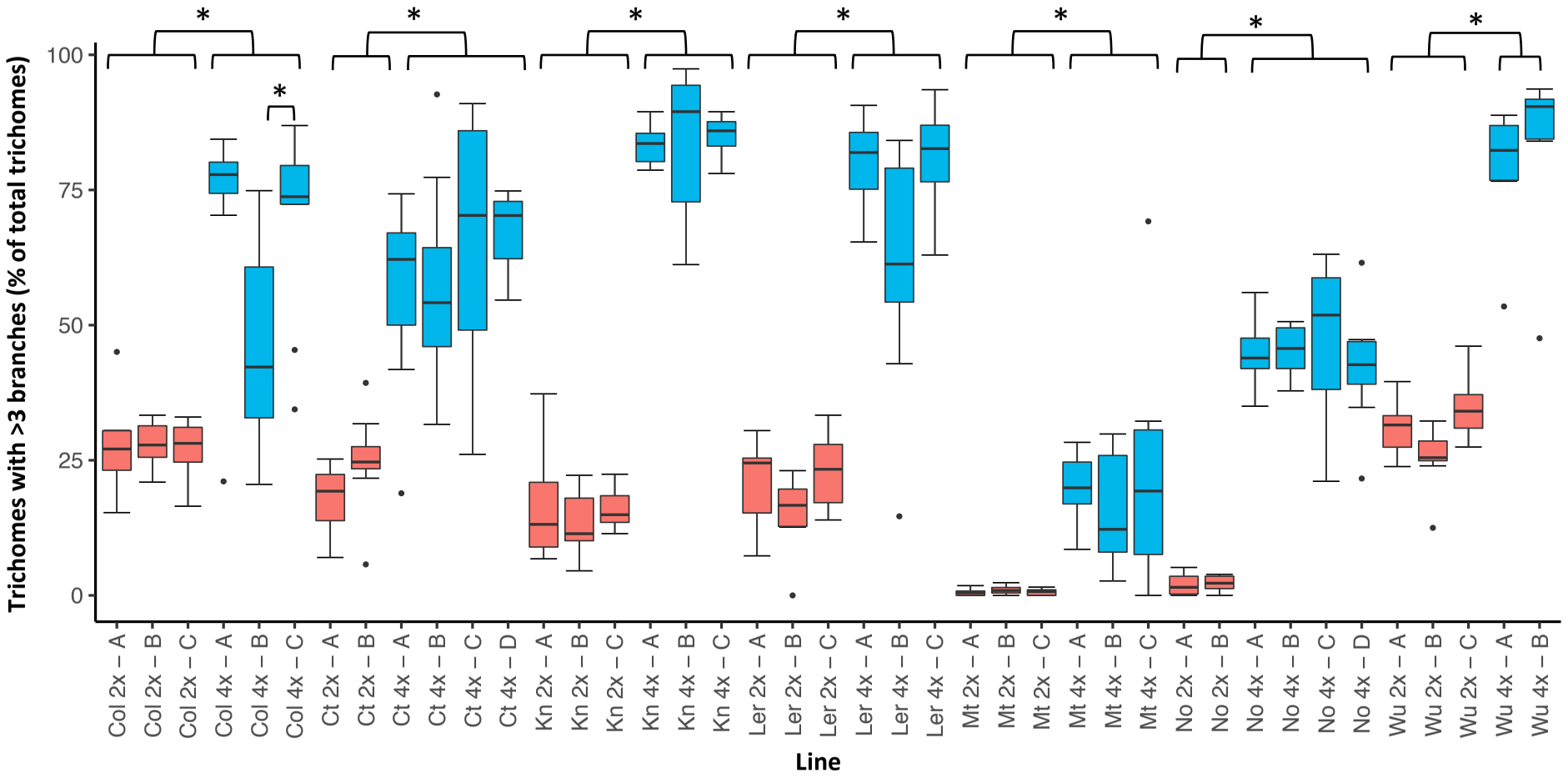


Figure 2.17 | Trichome branching data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers. Significance (*) : adjusted p-value < 0.05 using a Tukey's HSD test corrected for multiple comparisons.

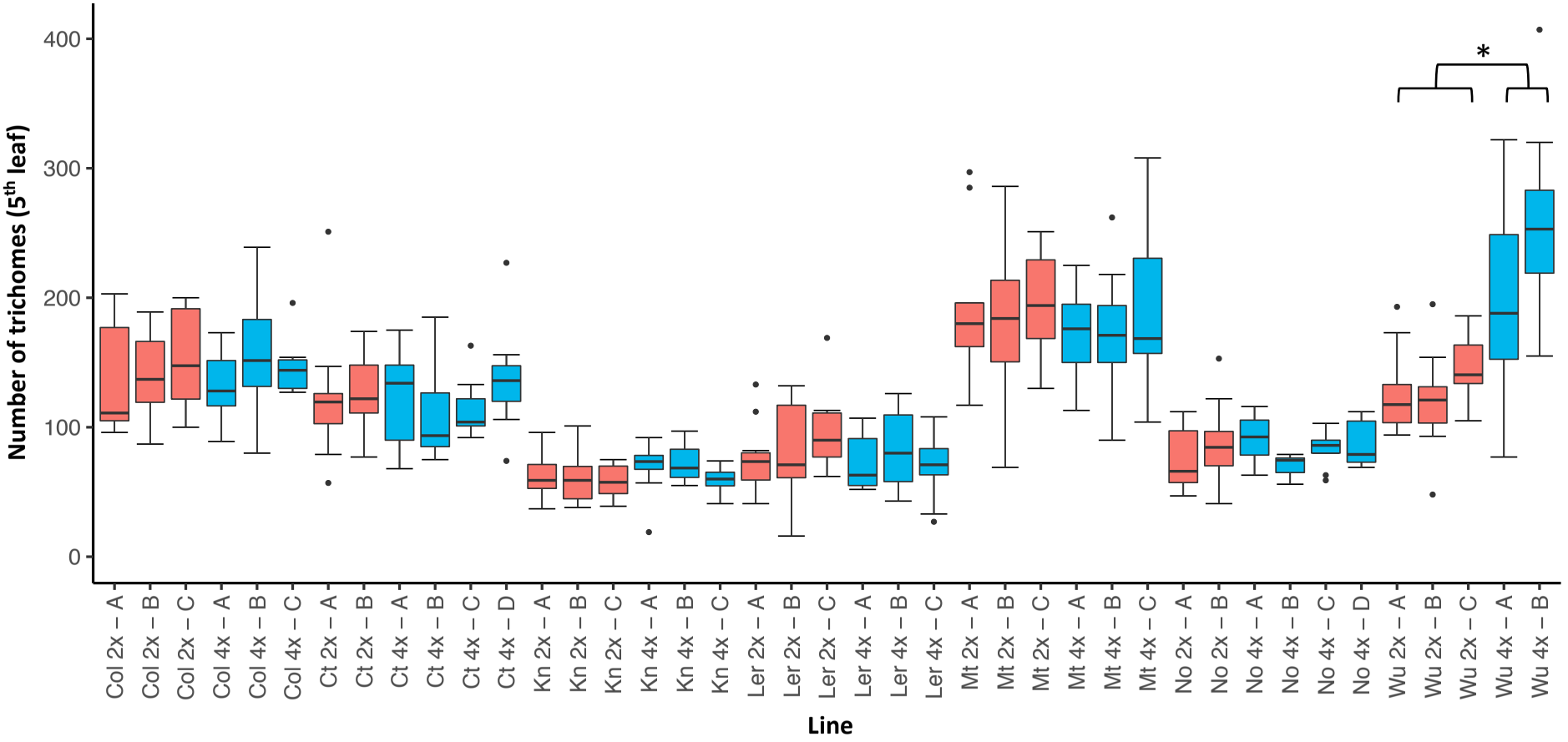


Figure 2.18 | Trichome number data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers. Significance (*) : adjusted p-value < 0.05 using a Tukey's HSD test corrected for multiple comparisons.

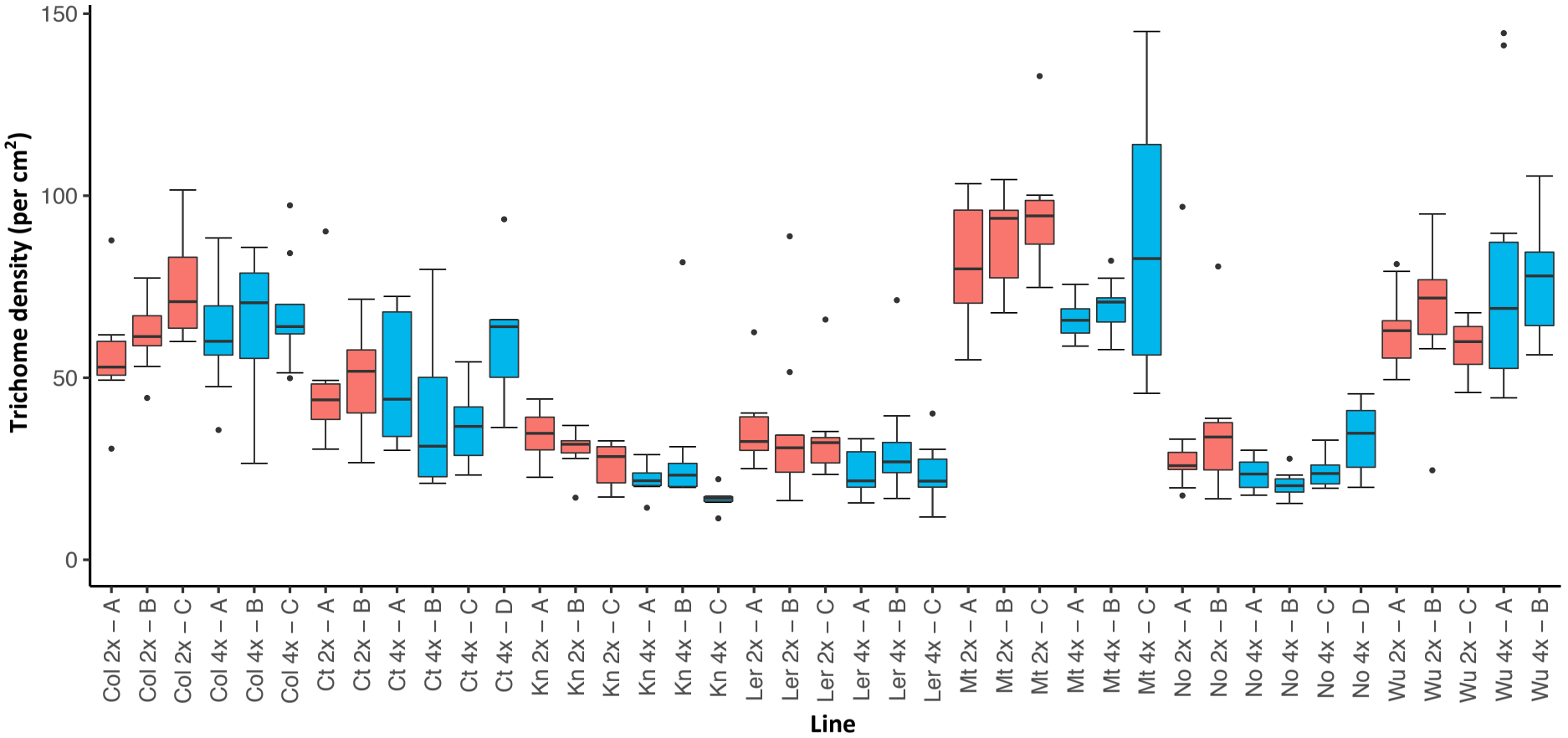


Figure 2.19 | Trichome density data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers.

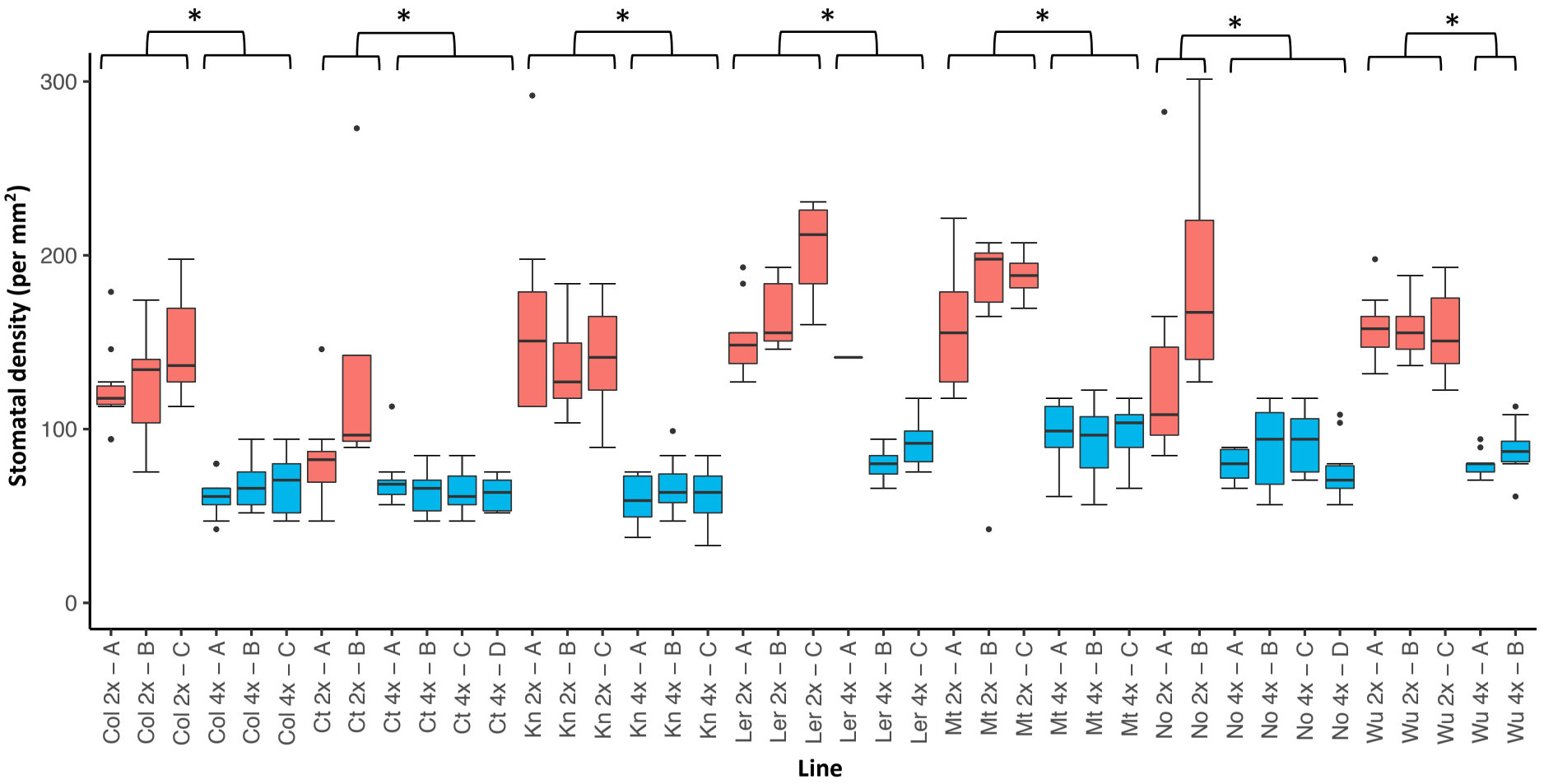


Figure 2.20 | Stomatal density data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers. Significance (*) : adjusted p-value < 0.05 using a Tukey's HSD test corrected for multiple comparisons.

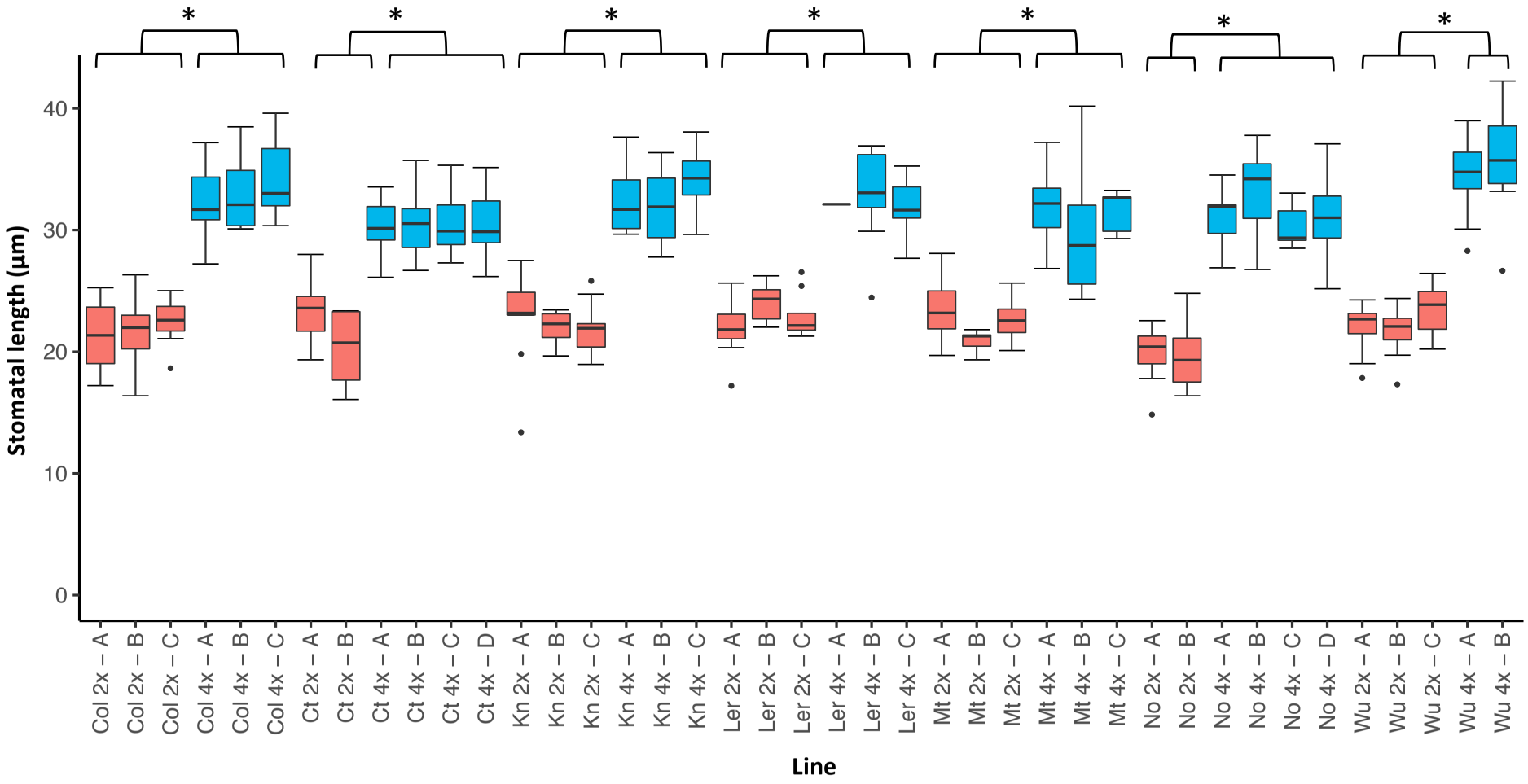


Figure 2.21 | Stomatal length data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers. Significance (*) : adjusted p-value < 0.05 using a Tukey's HSD test corrected for multiple comparisons.

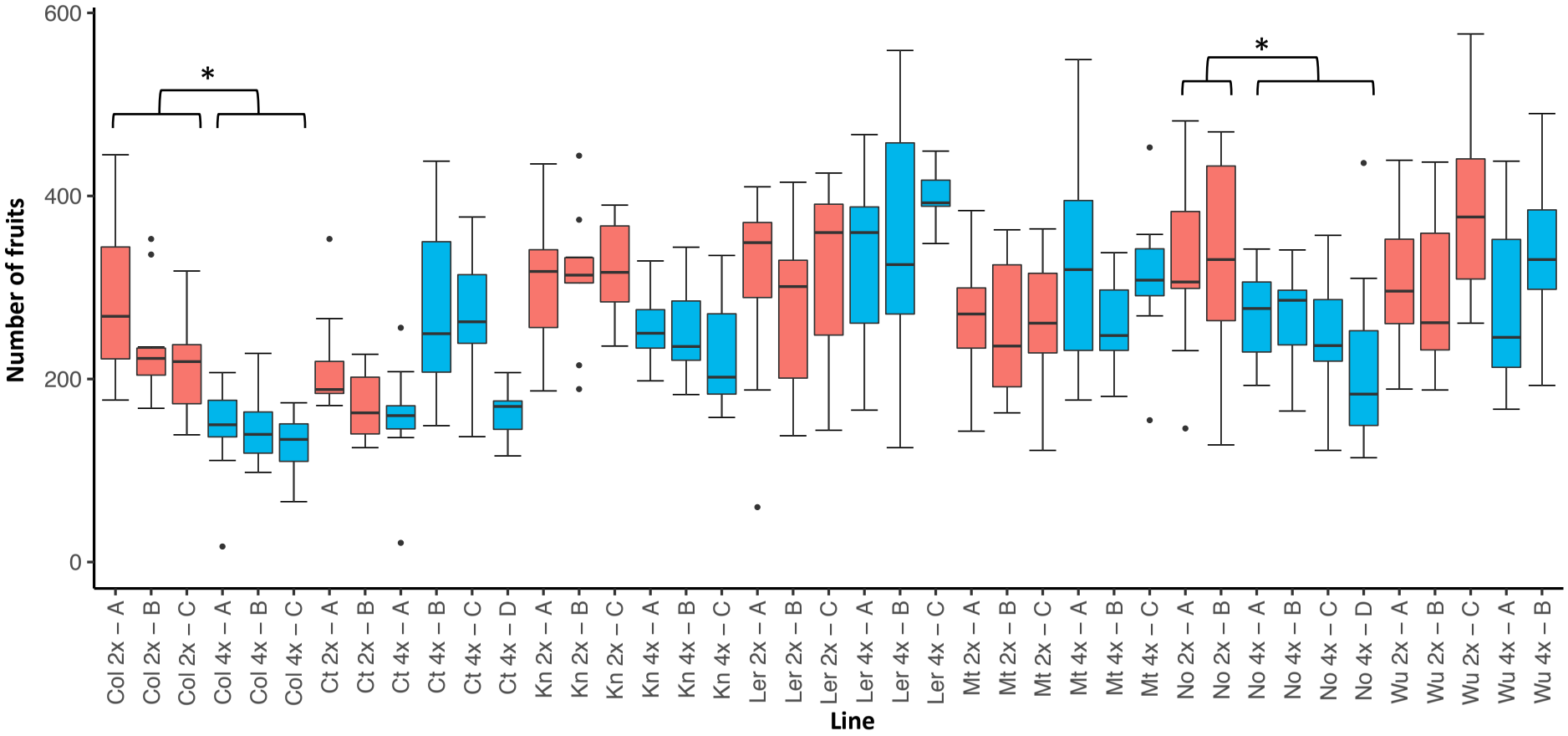


Figure 2.22 | Fruit number (fitness) data for colchicine treated diploids (red) and colchicine treated tetraploids (blue) for seven natural *A. thaliana* genotypes. X-axis labels indicate the genotype, ploidy level (2x or 4x) and the independently derived line (A, B, C or D). N=10 for each line. Boxes represent the inter-quartile range for each line; the horizontal bar within each box represents the median, the whiskers represent the maximum and minimum values; and black dots represent outliers. Significance (*) : adjusted p-value < 0.05 using a Tukey's HSD test corrected for multiple comparisons.

2.4 DISCUSSION

2.4.1 Colchicine has heritable effects on plant phenotype

A series of studies on the effects of colchicine in the late 1980's and early 1990's suggested that changes in colchicine-treated diploids were heritable and stable in ryegrass *Lolium perenne* (Francis & Jones, 1989) and cotton cultivar *Sicot 1* (Luckett, 1989) on various agronomic characteristics. Heritable changes in mesophyll cell plan, chloroplast numbers, and changes in chromosome behaviour at meiosis in terms of change in chiasma frequency and distribution were observed in diploid ryegrass and inbred ryegrass (Francis *et al.*, 1990; Hassan *et al.*, 1991; Hassan & Jones, 1994, 1995). However, the effects that colchicine may have when studying colchicine treated polyploid plants have often been overlooked.

Over the past decade, there have been several hundred studies that have looked into understanding the effects of polyploidy by artificially creating polyploids via colchicine treatment and comparing the neoautopolyploids to untreated diploids. When taking this approach, most polyploidy studies assume that colchicine itself has no effects on phenotype besides increasing ploidy. A few ploidy studies so far have specifically tested for the effects of colchicine on plant phenotype (Husband *et al.*, 2016; Münzbergová, 2017). Husband *et al.*, (2016) reported that colchicine led to smaller rosette diameters, decreased plant height and delayed flowering in the treatment generation (G_0) in *C. angustifolium,* but in contrast to our findings they observed that these effects were lost in the following generation after selfing (G_1). Münzbergová (2017), in concurrence with our data, showed that colchicine effects were seen on seed production and stomatal size in *Vicia cracca* plants two generations after colchicine treatment. While we observed that the deleterious effects of colchicine were most pronounced in the generation that was directly exposed to colchicine (G_0), we found that the phenotypic effects of colchicine, albeit subtle, were also seen in the third generation after treatment. Thus, the effects of colchicine independent of ploidy are heritable and need to be considered when interpreting phenotypic responses to ploidy in neoautopolyploid studies as the changes observed may not explicitly be due to chromosome doubling but could also be artefacts of colchicine exposure. Similar effects may also be observed with other chemicals commonly used to synthesise artificial polyploids like oryzalin and trifluralin, and similar studies that test for their heritable effects should be conducted and accounted for in comparative studies between ploidy

levels.

It is unclear whether colchicine has any mutagenic effects. Yu *et al.*, (2010) in their study using diploid and colchicine treated tetraploid *Arabidopsis thaliana* did not find any mutants via common mutagenic screening protocols. Lukens *et al.*, (2006) in their study using colchicine treated *Brassica* allopolyploids did not find any mutagenic effects of colchicine either. Szadkowski *et al.*, (2011), however, found that the patterns of genetic rearrangements in allopolyploid *Brassica napus* differed depending on their pathway of synthesis (unreduced gametes v/s somatic doubling using colchicine), suggesting that the behaviour of polyploids may be dependent on the pathway by which they were synthesised. Polyploid genomes may not always be perfect genetic multiples of their haploid genome, and they accumulate genetic rearrangements over time that may affect phenotype. Furthermore, the method used to generate polyploids may have an effect on the rearrangements, and consideration of the method should not be neglected. In our experiment, we used colchicine exposed diploid lines as controls (as opposed to untreated progenitor diploids) when comparing colchicine treated tetraploids to diploids to study phenotypic changes. We anticipate that, by using such an approach, any colchicine effects should affect both groups equally and any differences between colchicine treated diploids and colchicine treated tetraploids would be related to ploidy only.

2.4.2 Colchicine effects on plant phenotype are not likely to be stochastic

As colchicine was shown to have heritable effects on phenotype, it was important to understand if the effects differed for each event of colchicine treatment (i.e. stochastic) or consistent across all colchicine treated lines (i.e. reproducible). While we observed no significant variation between lines of the same genotype and ploidy in the Tukey's HSD (except for 1 comparison; Supplementary Data 7.3), suggesting that colchicine had no stochastic effects on plant phenotype. Pignatta *et al.*, (2010) found gene expression changes to be stochastic in three independently derived *Arabidopsis thaliana* tetraploid lineages. They harvested aerial parts of 4-week old plants and used microarray data to understand gene expression changes and found that the differentially expressed genes obtained between independently derived third generation (G_3) diploid and tetraploid pairs originating from the same treated parent (G_1) did not overlap. They used one genotype to study stochasticity, Col, which is the

same genotype for which we found one significant Tukey comparison for trichome branching – a leaf trait. As their plant material for RNA extraction also consisted of leaves, it may be possible that Col leaves exhibit stochastic gene expression differences between independently derived tetraploid lines, but gene expression differences may not translate to phenotype changes to the same extent. Moreover, as we used seven genotypes in our study, it may also be possible that only Col shows some phenotypic stochasticity and the other genotypes do not, suggesting that colchicine effects on phenotype, overall, may be reproducible for most genotypes. Given that some genotypes may show stochastic effects for certain traits (e.g. Col for trichome branching), in an ideal experiment it would be good practice to employ independent replication of the process of creation of the colchicine treated lines to establish a baseline variation in phenotype that colchicine is responsible for, so that the stochastic effects of colchicine can be unpicked from those of ploidy. However, this may not always be feasible due to the tedious nature of generating multiple independent lineages for multiple genotypes and requiring an increased sample size (by the same multiple) to capture minor stochastic variation that may or may not affect our interpretation of ploidy effects on phenotype.

2.4.3 Autotetraploids exhibit phenotypic differences in response to ploidy

If we were to assume that neoautotetraploids are perfect duplicates of their diploids, then, at a qualitative level, they would be isogenic. In this case, we would not expect to see any genetic diversity and consequently no phenotypic diversity in newly formed tetraploids unless other factors (genetic, epigenetic or biophysical) introduced some changes. However, altered phenotypes and ecological tolerance have commonly been reported in polyploids relative to their diploids (Parisod *et al.*, 2010; Bomblies & Madlung, 2014; Shi *et al.*, 2015; Gallagher *et al.*, 2016), the basis for which are not entirely understood. Traits like increased organ size in *Rosa rugosa* (Allum *et al.*, 2007), *Paulownia tomentosa* (Tang *et al.*, 2010), *Vitex agnuscastus* (Ari *et al.*, 2015), and *Escallonia rubra* (Denaeghel *et al.*, 2018); increased seed size in *Citrullus lanatus* (Zhang *et al.*, 2019a), increased trichome branching in *Arabidopsis thaliana* (Yu *et al.*, 2009); increased stomatal size in *Bletilla striata* (Li *et al.*, 2018), and *Arabidopsis thaliana* (Yu *et al.*, 2009); and decreased stomatal density in *Bletilla striata* (Li *et al.*, 2018) have often been reported as a common responses to increased ploidy in several species. Our data

also agrees with such size-related ploidy effects as we see increased trichome branching and larger stomata in tetraploids of all seven genotypes. These traits are thought to be directly correlated with genome size increase and appear to respond to ploidy increase independent of the genotype of the plant. However, we also observed significant genotype-by-ploidy interaction effects for these traits suggesting that although the responses may be universal, they are not uniform i.e. the extent of the response is dependent on the genotype.

Increase in DNA content of the nuclei can alter cell size and volume, which changes ratios of cytoplasm-to-membrane, inter and intracellular communication, the size and number of cell organelles and their interaction within the cell (Orr-Weaver 2015). While these changes seem plentiful to bring about significant changes in the cellular environment and plant phenotype, ploidy responses are not likely to be passively regulated by cell-size alone. Tsukaya (2008) showed that octoploid plants were much smaller than their diploid and tetraploid counterparts, suggesting that plant organ size is not passively regulated by ploidy. Studies on tetraploidized cell size mutant lines also revealed that increasing ploidy did not affect all lines equally (Tsukaya, 2013), suggesting that cell size itself is not passively regulated by ploidy. Quite often, the phenotypic effects of ploidy are discussed as a generalized response to increase in ploidy. However, it is possible that the observed phenotypic changes are the result of changes in the expression or regulation of particular genes. To understand such processes, it is important to characterise phenotypic traits in multiple genetic backgrounds.

2.4.4 Phenotypic responses to increased ploidy are genotype-dependent

The genotype-dependent relationship between phenotype and ploidy is particularly interesting and has been largely unexplored for other phenotypic traits like flowering time, rosette diameter and fitness. Furthermore, most ploidy studies in *Arabidopsis thaliana* have been limited to commonly used genotypes like Col, findings from which can't be extrapolated to a species-level interpretation of ploidy effects unless tested across multiple genotypes. Our study provides quantitative data across a range of phenotypes for a range of genetic backgrounds to aid our understanding of ploidy responses that do not appear to be regulated by nucleotype alone.

We observed significant genotype-by-ploidy interaction effects for all traits

studied, suggesting that for traits like stomatal size and trichome branching that show a consistent response to increased ploidy, there is genotype-dependant variation in the magnitude of the phenotypic change. Thus, nucleotype/gigas effects and genotype appear to work in conjunction to determine the ultimate polyploid phenotype. The understanding of underlying genetic mechanisms that regulate ploidy responses may prove vital to our understanding of polyploidy mechanisms.

Our data is in agreement with past findings that have also reported ploidy effects depending on the genetic background of the plant. Hias *et al.*, (2017) studied two genotypes of diploid and neoautotetraploid apples (*Malus domestica*) and showed that 15 out of 17 physiological and morphological traits studied were affected by ploidy. Of the 15 traits, 10 exhibited common responses in both genotypes and five exhibited genotype-specific responses to increased ploidy. In another study they (Hias *et al.*, 2018) used three genotypes of *Malus domestica* and showed that the degree of tolerance of neoautotetraploids to fungal infection (apple scab caused by *Venturia inaequalis*) in comparison to diploids was dependent on genotype. Species-specific responses to increased ploidy have been reported in neoautotetraploid *Solanum* species (Aversano *et al.*, 2015; Fasano *et al.*, 2016) and *Escallonia* species (Denaeghel *et al.*, 2018) where the effects seen in neoautotetraploids was dependent on the diploid parent from which they were derived. Eliášová & Münzbergová, (2014) observed inter-population variation in phenotype in natural populations of autotetraploid *Vicia cracca.* Findings from these studies strongly suggest that for several phenotypic traits, the response to increased ploidy is not consistent across the genotypes or species studied and is often largely dependent on the genetic background of the plant.

Most quantitative traits are polygenic and often have a complex network of genes interacting with one another to determine the final phenotype. Failure to account for variation in the response to ploidy that may be caused by genetic differences excludes the possibility of understanding how the doubling of specific genes or gene combinations elicit specific responses. Polyploids have been known to undergo extensive chromosomal rearrangements (Weiss & Maluszynska, 2000), alternative splicing (Zhou *et al.*, 2011) and epigenetic modifications (Osborn *et al.*, 2003; Leitch & Leitch, 2008). Lukens *et al.*, (2006) used RFLP (Restricted Fragment Length Polymorphism) analysis and found that genetic changes were rare (<0.2% per locus) but epigenetic changes in CpG methylation were relatively frequent (7% of the loci are

affected) in resynthesized *Brassica napus*. Interestingly, their data showed that methylation changes occurred at specific loci and the findings were consistent across the population of neoallopolyploid *Brassica napus.* This reiterates the need to study multiple polyploid genotypes as genomic rearrangements and epigenetic changes that eventually give rise to phenotypic changes in polyploids may not be stochastic but in fact dependent on the specific gene loci or gene combinations in the plant. Although epigenetic variation can contribute to and regulate phenotypic and transcriptomic variation in polyploids (Comai *et al.*, 2000; Wang, 2004; Song & Chen, 2015; Ding & Chen, 2018), it is much more frequent in allopolyploids than autopolyploids. This could, in part, be the reason why more pronounced phenotypic and transcriptomic alterations are reported in allopolyploids in comparison to autopolyploids (Wang *et al.*, 2006; Spoelhof *et al.*, 2017). Nonetheless, there is some similar evidence for the absence of polymorphism but presence of epigenetic modifications autopolyploids as well (Stupar *et al.*, 2007; Aversano *et al.*, 2013). Changes in alternative splicing patterns have been reported in neoallopolyploid *Brassica napus* in which two independently derived resynthesized lineages showed the same alternative splicing pattern between the lineages but a 26-30% change in alternative splicing was observed when compared to the parents (*Brassica rapa* and *Brassica oleracea*), thus suggesting that alternative splicing is not a random process (Zhou *et al.*, 2011).

In summary, phenotypic changes observed in polyploids are regulated by complex mechanisms of nucleotype and associated effects, epigenetic effects, and alternative splicing that may work in isolation or in conjunction to alter gene regulatory networks that govern several quantitative traits. These mechanisms are likely to be genotype dependent as phenotypic responses to increased ploidy are found to vary between different genotypes. Thus, to be able to extrapolate genotype-level findings to generalized polyploid phenomena, it is important to study quantitative traits in multiple genotypes and species. Finally, for the accuracy of autotetraploid v/s diploid comparisons and correct interpretation of ploidy effects, it is vital to take into consideration the changes that may be caused by the process of somatic doubling (e.g. colchicine treatment).

Chapter 3 : NATURAL GENETIC VARIATION IN GENE EXPRESSION RESPONSES TO INCREASED PLOIDY IN *ARABIDOPSIS THALIANA*

Sarah Chordekar, Atahualpa C. Morales, Araxi O. Urrutia and Paula X. Kover

Author Contributions:

SC: Performed the experiments, analysed the data, wrote the draft

ACM, AOU: Advice on bioinformatic data analysis

PXK: Supervised, reviewed the draft.

3.1 INTRODUCTION

Polyploidy is very common in plants, and it may arise by two main paths: whole genome duplication (autopolyploidy), or the combination of two different genomes through hybridization (allopolyploidy) (Comai, 2005). Polyploidy is often associated with significant changes in phenotype, including increased vigour and environmental tolerance (McIntyre & Strauss, 2017; Godfree *et al.*, 2017; Klatt *et al.*, 2018; Čertner *et al.*, 2019; Welles & Ellstrand, 2019). The molecular mechanisms that bring about such changes in phenotype are expected to occur either through DNA sequence changes (mutation, recombination, hybridization or rearrangements) that encode expressed RNA or proteins; or through changes in the expression of the DNA sequence (transcriptional, translational and/or epigenetic) that regulate gene expression and consequently alter the abundance of a given protein, and the temporal or spatial variation in its expression.

When comparing the transcriptome data of paleopolyploids to their respective diploids, it is observed that some genes are preferentially retained (genes involved in transcription and signal transduction) whereas some are preferentially lost (genes coding for organellar proteins and those involved in DNA repair mechanism), thus suggesting that gene loss or retention over evolutionary time is not a random process (Blanc & Wolfe, 2004a; Seoighe & Gehring, 2004; Kaltenegger *et al.*, 2018). However, in recent polyploids (e.g. synthetic polyploids), one would not expect to see such DNA sequence changes immediately after whole genome duplication, and thus, any phenotypic changes observed between diploids and neo-polyploids are likely to be due to gene or protein expression changes. When the entire genome is duplicated, the simplest assumption is that the increase in copy number will affect the expression of all genes equally, and no change in relative gene expression should be detected (Doyle & Coate, 2019). However, it is possible that the relationship between the number of DNA copies and their expression is neither linear (i.e. proportional), nor constant for all coding genes.

Gene expression studies in *A. suecica* (an allotetraploid hybrid of *Arabidopsis thaliana* and *Arabidopsis arenosa*) revealed that 5.6% of the genes (1469 genes) diverged in expression from the midparent value in two independently derived synthetic lines (Wang *et al.*, 2006). The increased allelic variation brought about by the merger of

two genomes provides opportunities for novel gene combinations and is often thought to be one of the reasons for such gene expression changes. Studies have also shown that the homeologous loci in allopolyploids are unequally expressed (Wang *et al.*, 2006; Yoo *et al.*, 2013; Xiang *et al.*, 2019). Thus, after an event of whole genome duplication through hybridization (allopolyploidy) a broadening of the phenotypic diversity expressed is expected, which via natural selection can lead to the fixation of new phenotypes (Shi *et al.*, 2015; Gallagher *et al.*, 2016; Shimizu-Inatsugi *et al.*, 2017; Hu & Wendel, 2019), or the development of new agronomical varieties with higher yield for example (Renny-Byfield & Wendel, 2014).

Whilst in nature there is a cascade of variation from fully homozygous autopolyploids (from spontaneous whole genome duplication) through to autopolyploids from divergent populations to allopolyploids (Doyle & Sherman-Broyles, 2017), in order to specifically study the effect of genome doubling *per se* on transcriptomic changes and isolate them from any changes due to hybridization or natural selection, we focus on newly generated synthetic autopolyploids.

In contrast to allopolyploids, recent autopolyploids have been shown to exhibit relatively minor gene expression differences between diploids and tetraploids. For example, only 88 differentially expressed genes (~0.3% of the total genes) were observed between diploid and autotetraploid *Arabidopsis thaliana* using spotted oligo-gene microarrays (Wang *et al.*, 2006). In another microarray study, a total of over 200 differentially expressed genes (~1% of the total genes) were observed between diploids and neotetraploids *Citrus limonia* (Allario *et al.*, 2011). Few other autopolyploid studies have found comparable findings in species like *Solanum phureja* (Stupar *et al.*, 2007), *Brassica oleracea* (Albertin *et al.*, 2005), *Zea mays* (Riddle *et al.*, 2010), *Chrysanthemum lavandulifolium* (Gao *et al.*, 2016) and *Manihot esculenta* (Yin *et al.*, 2018), where little gene expression or protein expression differences were reported in response to increased ploidy.

Based on past literature, it could be suggested that transcriptome alterations in autopolyploids are negligible. However, with newer methods (like RNA-Seq, single-cell

sequencing and ERCC RNA spike-ins[5]) and improved statistical tools to analyse transcriptome data, this question has been revisited in some species and have yielded improvements in the ability to detect differentially expressed genes. For example, when using microarrays, only a small number of differentially expressed genes (between 0.3%-0.9% depending on the genotype studied) were reported in autotetraploid *Arabidopsis thaliana* (Wang *et al.*, 2006; Pignatta *et al.*, 2010; Yu *et al.*, 2010). In comparison, when using RNA-Seq, a relatively larger number of differentially expressed genes were detected – from 3% (743 genes) in *Arabidopsis thalina* (Zhang *et al.*, 2019b) to 28% in watermelon species *Citrullus lanatus* (Saminathan *et al.*, 2015) where 6650 genes were found to be differentially expressed between diploids and tetraploids. This reflects that the lack of transcriptomic changes reported in neoautotetraploids in past studies may not necessarily be evidence of absence of gene expression differences, but rather be due to limitations in the sensitivity of microarrays in detecting subtle changes or in separating low-level gene expression changes from experimental noise (Pignatta *et al.*, 2010). Thus, accruing more gene expression data using newer sequencing technologies are needed in order to understand genetic consequences of autopolyploidy as they are more sensitive in detecting transcriptome alterations (Wilhelm & Landry, 2009), novel transcripts and alternative splicing (Saminathan *et al.*, 2015; Li *et al.*, 2016).

In addition to analysing neo-autopolyploid transcriptomes with advanced sequencing approaches, it may be important to incorporate the use of multiple genetic backgrounds before generalizing the changes seen in one genotype as a universal response of the species to increased ploidy. It has been previously noted that there is genetic variation in phenotypic traits in response to increased ploidy (chapter 2 and references therein). Thus, it forms a reasonable hypothesis to test whether there exists any variation in gene expression changes between different genotypes in response to increased ploidy. If the effects of ploidy are independent of genotype, then any gene expression changes observed between diploids and tetraploids should be consistent across all genotypes. Any variation in gene expression between the genotypes would suggest the presence of allelic variation in response to increased ploidy. Such allelic variation could be the potential factor underlying phenotypic variation seen in different

[5] ERCC RNA spike-ins are synthetic RNA standards developed by the External RNA Controls Consortium (ERCC) that facilitate measurement of absolute gene expression levels in transcriptome studies (Pine *et al.*, 2016)

autopolyploid genotypes that could act by modulating gene regulatory networks, generating different splice variants, or modifying epigenetic programming.

Past studies on understanding the relationship between ploidal levels and gene expression have seldom gone beyond diploid and tetraploid comparisons (but see Guo *et al.*, 1996; Stupar *et al.*, 2007; Yao *et al.*, 2011; Robinson *et al.*, 2018). Using a third ploidal level (e.g. octoploids) can provide an opportunity to learn about the trends in gene expression changes with increasing ploidy, which would not otherwise be evident in comparisons between two ploidy levels. If gene expression changes between diploids, tetraploids and octoploids are linear, it may be possible to detect genes that express subtle changes in response to ploidy (in diploid to tetraploid comparisons) that are generally not large enough to pass through the stringent statistical thresholds of programs that detect differentially expressed genes. By adding an additional level, diploids can be compared to octoploids and the expectation is that if limitations in detection are due to low-level gene expression changes, then in diploid to octoploid comparisons, the differences should be scaled making it easier to unpick gene expression changes due to increased ploidy from background noise. However, it is also possible that gene expression changes are not linear (1:1) between diploids, tetraploids and octoploids, or even in the same direction (i.e. an increase in gene expression in tetraploids relative to diploids may not necessarily imply an increase in octoploid gene expression relative to tetraploids or diploids). In this case, it is interesting to understand whether or not the relative abundance of transcripts remains constant across the ploidies (this would be expected if all genes at a ploidy level increase in expression by the same proportion, thereby keeping the gene expression ratio between the genes constant).

Here, we used *Arabidopsis thaliana* as a model to study gene expression changes using RNA-Seq in a polyploidy series (diploid, tetraploid and octoploid lines) for two genetic backgrounds. As the process of developing neo-autopolyploid lines involved treating seedlings with colchicine, we used colchicine treated diploids as controls in addition to untreated diploids to isolate the effects of ploidy *per se* from any confounding effects of colchicine treatment (the importance of this is highlighted in chapter 2). Furthermore, all lines used in this experiment were also phenotyped for various traits in order to link gene expression changes with the phenotypes observed.

Specifically, we addressed (i) whether there were significant gene expression changes with increased ploidy (ii) if the gene expression changes were genotype-dependent (iii) if the changes between diploids, tetraploids and octoploids were linear (iv) did the differentially expressed genes explain the phenotypic changes observed.

This is the first study to use RNA-Seq to study gene expression changes in multiple genotypes and across three ploidy levels in *Arabidopsis thaliana*. Furthermore, the inclusion of colchicine treated diploids as controls makes it a robust model to efficiently isolate colchicine, genotype and ploidy effects on gene expression.

3.2 MATERIALS AND METHODS

3.2.1 Generating polyploid lines

Diploid seeds for Ler (CS20) and Sf (CS6857), two natural *A. thaliana* genotypes, were originally obtained from the Arabidopsis Biological Resource Centre (ABRC, USA) and maintained in the lab for a few generations through selfing. Seeds from these lines, which were used for generating the polyploid lines used in this experiment, will be referred to as $Ler2x_{-untreated}$ and $Sf2x_{-untreated}$ hereafter.

Tetraploid and octoploid lines were generated by treating seedlings with colchicine as described by Yu *et al.,* (2009). Briefly, 50-80 seeds from each progenitor line were grown in individual 5cm diameter pots containing Levington® F2+S compost for two weeks. The shoot apical meristems of two-week old seedlings were treated with 10 µl 0.1% colchicine solution (Sigma-Aldrich®) and grown to maturity. All plants that survived the treatment were allowed to self for one generation. A single seed from each colchicine treated plant was grown and assessed for ploidy using flow cytometry (as described in Chapter 2). All plants determined to be diploid, tetraploid and octoploid from each of the genotypes after colchicine treatment were selfed for two more generations. This led to the production of the eight experimental lines used here: $Ler2x_{-untreated}$, $Ler2x_{-colchicine}$, $Ler4x_{-colchicine}$, $Ler8x_{-colchicine}$, $Sf2x_{-untreated}$, $Sf2x_{-colchicine}$, $Sf4x_{-colchicine}$ and $Sf8x_{-colchicine}$. The plants that had undergone colchicine treatment but remained diploid ($2x_{colchicine}$) provide a control for the effect of colchicine, independent of ploidy (as highlighted in chapter 2).

3.2.2 Experimental design

To determine the effect of whole genome duplication on gene expression and phenotype, we grew plants from the eight lines described above under identical

conditions. In the first experiment, 3rd generation seeds following colchicine treatment (G_3) were grown to collect leaf tissue for RNA extraction. To ensure that plants were not disturbed prior to tissue collection for RNA extraction, we only collected data for flowering day and post flowering phenotypes: rosette diameter, seed size, seed number, petal size and pollen size (described in section 3.2.3). For the second experiment, mature seeds produced by plants in the first experiment (i.e. G_4 seeds) were grown in conditions identical to the first experiment, but seedlings were destructively sampled to analyse stomatal size, trichome number and leaf area (as described in section 3.2.3).

In each experiment, seeds from each line were sown in 20 replicate 5cm pots containing Levington® F2+S compost (for a total of 160 pots) and stratified at 4°C in the dark for four days. Pots were then randomly distributed among 10 trays and placed in a growth chamber under a variable programme designed to simulate natural conditions, as described in Scarcelli *et al.,* (2007). Briefly, plants start growing in autumn-like conditions where temperatures gradually decrease and the day length shortens, transitioning to a cold winter-like phase, and finally transitioning to spring and summer-like conditions where the temperature gradually increases to 21°C and the day length increases to 16 hours as depicted in Figure 3.1. Five seeds were sown in each pot and

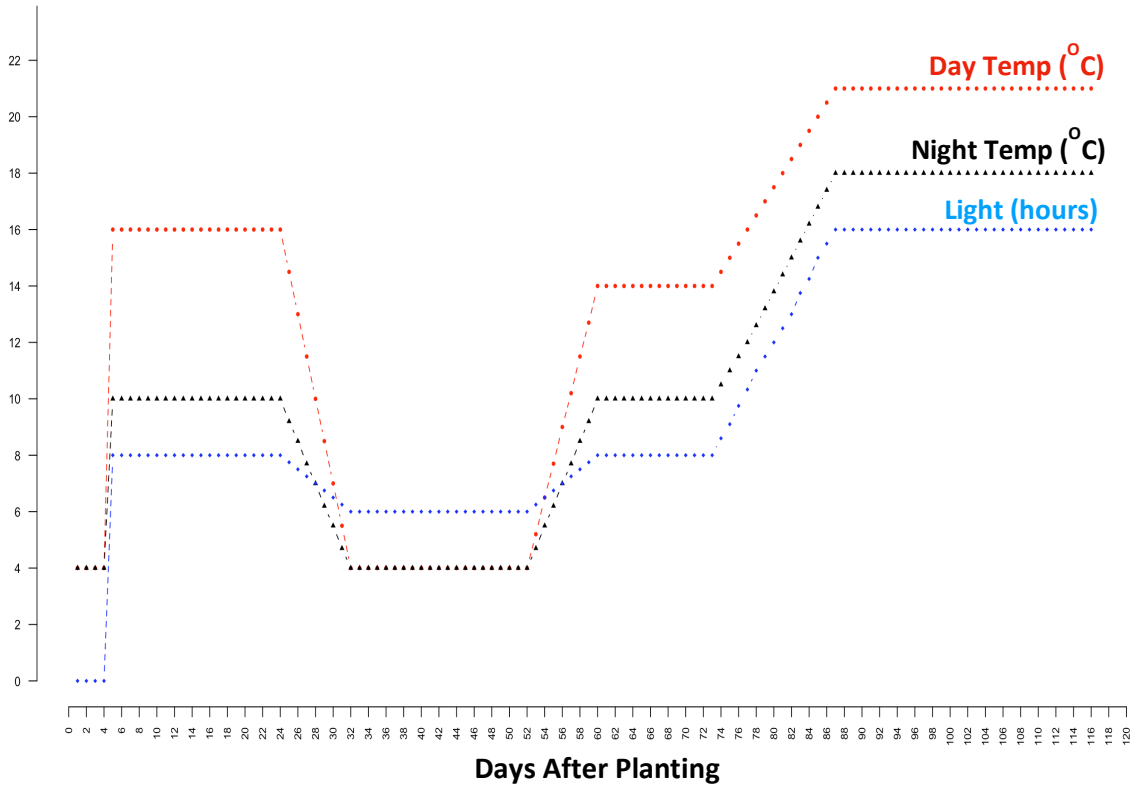


Figure 3.1 | Light and temperature conditions used in the experiment. A dynamic light and temperature regime were set to artificially simulate autumn, winter, spring and summer seasons within the growth chamber to provide ecologically relevant environmental cues for plant growth. The blue line represents the photoperiod and the red and black lines represent the day and night temperatures respectively on each day after planting.

their germination was monitored daily. When the first true pair of leaves emerged, all pots were thinned down to retain one seedling per pot. Plants were bottom watered as needed, and trays rotated every week to minimise microenvironmental variation within the growth chamber. Plants were fertilised every three weeks (250mL of a 1.66g/L of Vitax Soluble Balanced Multipurpose Vitafeed 111 per tray).

3.2.3 Estimating the effects of ploidy on phenotypic changes

3.2.3.1 Flowering time

Flowering time was defined as the time between the germination day and the appearance of first white petals. Pots were monitored daily and the day when the cotyledons were first open in each plant (germination day), as well as day when the first open flower was observed in each plant (flowering day) was recorded and the difference between the two was calculated as the flowering time.

3.2.3.2 Rosette diameter

After the plants had flowered, and the leaf tissue for RNA extraction had been collected (described in section 3.2.4.1), the diameter of the widest part of the rosette was measured using a ruler.

3.2.3.3 Pollen size

Mature flowers were collected from the primary inflorescence of five to eight plants of each line (depending on the availability of healthy surviving replicates per line) and kept separately. To estimate pollen size, the anthers of one flower per plant were dabbed on a glass slide to release pollen. The pollen grains were hydrated for 60 seconds by adding a drop of water to the slide. The region was then gently covered with a cover slip and pictures of 10-30 hydrated pollen per flower were taken under 400x

magnification of a Confocal Microscope (Nikon) coupled to a Digital Sight DS- U1 colour camera (Nikon) using NIS Elements-F software. The average pollen area for each flower was later measured using imageJ (Schneider *et al.*, 2012), using a stage micrometer as a reference scale.

3.2.3.4 Petal size

To measure the average petal area, one flower was collected from five to eight plants of each line (depending on the availability of healthy surviving replicates per line) and was immediately submerged in an eppendorf tube containing ethanol to prevent shrivelling. The flowers were then carefully dissected in a drop of water on a glass slide under a dissecting microscope to separate the petals from the other floral organs. The petals (four petals per flower) were then covered with a coverslip and the edges were sealed using nail varnish to prevent the petals from dehydrating. The petals were imaged against a dark background under a dissecting microscope coupled to a camera and their surface area was measured using imageJ (Schneider *et al.*, 2012). The average area of four petals from each flower was calculated for each plant replicate.

3.2.3.5 Seed size and seed number

Once the fruits had matured and turned brown, five fruits were collected from the lower to middle region of the primary axis from five plants for each of the lines (for a total of 5 fruits X 5 plants X 8 lines = 200 fruits). Seeds from the five fruits from each plant replicate were pooled and imaged under 14x magnification using a macro lens (Olloclip®, USA) coupled to an iPhone camera (Apple, USA). These images were later processed using ImageJ (Schneider *et al.*, 2012) to determine the average seed area (seed size) and average number of seeds per fruit (seed number).

3.2.3.6 Stomatal length

Five seedlings per line from the second experiment were collected when the first true pair of leaves began to emerge to ensure that the cotyledons had fully expanded. The cotyledons of each seedling were clipped, placed on a glass slide, and the abaxial surface was covered with a thin layer of nail varnish and allowed to dry. The nail varnish was then peeled gently with the help of a clear piece of Sellotape and stuck onto a fresh labelled glass slide. The central region of the cotyledon imprints were imaged at 400x

magnification using a Nikon Confocal Microscope coupled to a Digital Sight DS- U1 colour camera (Nikon) using NIS Elements-F software. The stomatal diameter of five stomata were measured per cotyledon using imageJ (Schneider *et al.*, 2012) calibrated using a stage micrometer as reference scale (5 measurements x 5 plants x 8 lines). The average of these five measurements was used to estimate stomatal diameter per cotyledon per replicate plant. In cases where the stomatal imprint was unclear due to uneven spreading of the nail varnish layer at a microscopic level, the imprint of the second cotyledon from the pair collected was imaged. The remaining seedlings (i.e. the ones not destructed for stomatal imprints) were allowed to grow and were thinned to one healthy seedling per pot when the first true pair of leaves had fully emerged.

3.2.3.7 Leaf size and leaf shape ratio

After the plants had bolted (i.e. appearance of the first floral buds at the centre of the rosette), the largest leaf was clipped from five random plants per line and preserved overnight in a petri-dish containing absolute ethanol. The decolourised leaf was then transferred to water in order to facilitate softening and flattening of curled leaf edges to get a flatter surface to image. A reference scale was added to each image, and using ImageJ (Schneider *et al.*, 2012) we calculated the leaf surface area and the leaf shape ratio (length to width ratio).

3.2.3.8 Trichome branching and density

After imaging each leaf for size and shape ratio, their adaxial surface was observed under a dissecting microscope to count the total number of trichomes on the entire leaf. The number of branches in each trichome were counted for all trichomes on the entire leaf to calculate the percentage of 2-branched, 3-branched, 4-branched and ≥5-branched trichomes present on each leaf. To aid visualization of the translucent trichomes, an LED light source was placed perpendicular to the leaf surface to allow the trichomes to reflect the light thus appearing clearly visible and making the analysis of trichome number and branching patterns more accurate. Using the estimated leaf area from the images (described above), we also calculated trichome density (trichomes per cm^2) of the leaf.

3.2.3.9 Data analysis of phenotypic traits

Statistical analysis was done using the software R version 3.5.1 (R Core Team, 2018). Each phenotypic trait was analysed independently using the following nested fixed effects linear model: "*Trait ~ Genotype + Ploidy + Genotype:Ploidy + Ploidy/Treatment*" where Genotype:Ploidy is the genotype by ploidy interaction term and Ploidy/Treatment is treatment nested within ploidy. We use a nested model for the treatment term as only one ploidy level (2x) has multiple treatment groups ($2x_{\text{-untreated}}$ and $2x_{\text{-colchicine}}$). All plots were generated using the ggplot2 R package (Wickham 2016).

3.2.4 Estimating the effects of ploidy on gene expression

3.2.4.1 Sample collection and RNA extraction

Three disks of 8mm diameter leaf tissue were collected from the 9th or 10th leaf from each of three plants of each of the eight lines one day after the first flower opened (to ensure that all plants were at the same stage of development when leaf tissue was collected). The tissue was immediately submerged in 1mL Tri-Reagent (Sigma-Aldrich) and stored at −80°C to prevent RNA degradation. Tissue samples were lysed in Tri-Reagent (Sigma-Aldrich) with 5mm stainless steel beads (Qiagen) using a TissueLyser (Qiagen). 200µL of chloroform was added to the homogenate and vortexed vigorously for 30 seconds and centrifuged for 15 minutes at 12000g at 4°C. The aqueous phase was transferred to a new tube and RNA was precipitated by adding 0.5x volume of ice cold 100% ethanol. This solution was then applied and spun through an RNeasy mini spin column (Qiagen) in two steps of ~500 µL each. An on-column DNase treatment was performed using the RNase free DNAse kit (Qiagen) according to manufacturer's protocol. The column was then washed and purified using the RNeasy Plant Mini Kit protocol (Qiagen) and RNA was eluted in 30 µL of RNAse-free water. RNA quality was monitored by running the samples on 1% agarose gels. All samples used for sequencing (described below) showed two intact rRNA bands with minimal smearing, and no evidence of DNA contamination. RNA samples were quantified using a Qubit fluorometer (Invitrogen) and stored at −80°C until further processing.

3.2.4.2 Library construction and Illumina sequencing

Sequencing libraries were generated using TruSeq stranded mRNA library prep kit (Illumina) at the IBERS Translational Genomics facility, Aberystwyth University, UK. All 24 samples (3 biological replicates x 8 lines) were multiplexed and sequenced on an

Illumina Hiseq2500 platform to generate 126bp paired end reads. To get a sequencing depth of approximately 20 million read pairs per sample (i.e. a minimum of 480 million read pairs totally), the samples were sequenced over three flow cell lanes. The three forward and reverse FASTA files produced by the three flow cell lanes for each sample were concatenated into a single forward and reverse file respectively for each sample before further processing.

3.2.4.3 RNAseq data analysis

3.2.4.3.1 Quality control and trimming

In total, we obtained 565.8 million pairs of reads, and an average of 23.6 million per sample. Raw reads were checked for overall quality using FastQC v0.10.1 (Andrews 2010). We used Trimmomatic v0.32 (Bolger *et al.*, 2014) to remove Illumina adaptors, bases of poor quality (Q<5) at the beginning and the end of the reads, and reads smaller than 60 base pairs (LEADING:5 TRAILING:5 MINLEN:60).

3.2.4.3.2 Quantification of read counts for each gene

The number of reads per gene in each sample were quantified using the software Kallisto v0.44.0 (Bray *et al.*, 2016); which quantifies transcript abundance from high throughput sequencing reads by a pseudoalignment approach. The cDNA reference genome for Arabidopsis (Araport11_genes.201606.cdna.fasta.gz) was downloaded from Araport (www.araport.org; accessed 01/05/2018). We then used the R package Tximport (Soneson *et al.*, 2015) to consolidate the read count data generated by Kallisto for each of the 24 samples into a single table.

3.2.4.3.3 Filtering and normalization of read counts

Initially, very lowly expressed genes (defined as genes with <1 count per million in 3 or more samples across all samples) were filtered out to reduce sample noise (which should increase the sensitivity of determining differentially expressed genes, (Risso *et al.*, 2011)). The filtered count table was then normalized using the TMM normalization method using the 'calcNormFactors' function in the EdgeR v3.22.2 (Robinson *et al.*, 2010; McCarthy *et al.*, 2012) R package to account for compositional differences between the libraries. Normalized gene count tables were prepared independently for: All samples (24 samples), Ler samples (12 samples), Sf samples (12 samples), Ler$_{\text{-colchicine}}$

samples (9 samples), and $Sf_{-colchicine}$ samples (9 samples) and used as needed for further analyses.

3.2.4.3.4 Exploratory data analysis

Dimensionality reduction analysis was performed on the normalized gene count data to visualise clustering patterns in the dataset using multidimensional scaling (MDS) plots constructed using the 'plotMDS' function in EdgeR v3.22.2 (Robinson *et al.*, 2010; McCarthy *et al.*, 2012) and principal component analysis (PCA) plots constructed using the 'prcomp' function in R.

3.2.4.3.5 Regression analysis

Normalized gene expression data for $Ler_{-colchicine}$ and $Sf_{-colchicine}$ were used to perform linear regression analyses for ploidy pairs (2x v/s 4x; 4x v/s 8x; 2x v/s 8x) to determine if there is any relationship between the expression levels for each ploidy pair. The coefficient of determination (R^2) was calculated to evaluate the fit of data points around the regression line and estimate what proportion of the variance in gene expression could be explained by the ploidy level. If there is no relative change in gene expression between the ploidy pairs, the slope of the regression line would be expected to be unity (m=1) and the intercept zero (c=0). To determine the extent to which the regression line deviates from unity, the 'smatr' R package function 'sma' was used as follows: s*ma(y ~ x -1, log="xy", data=data, slope.test=1).* This function tests if the slope of the regression line passing through origin is significantly different from the null hypothesis of slope=1.

3.2.4.3.6 Chromosomal location of gene expression changes

Fold changes in normalized gene expression between ploidy pairs were mapped to the chromosomal location of the genes using the annotation for Arabidopsis (Araport11_GFF3_genes_transposons.201606.gtf) downloaded from Araport (www.araport.org; accessed 23/03/2018) and plotted using ggplot2 v3.2.0 (Wickham 2016) in R.

3.2.4.3.7 Directionality of gene expression changes

To understand the nature in terms of the linearity of gene expression changes across the three ploidies, the expression data for each gene was categorised into the following groups : (1) No change in expression between 2x and 4x, and no change in expression between 4x and 8x; (2) Increase/decrease in expression between 2x and 4x,

and no change in expression between 4x and 8x; (3) No change in expression between 2x and 4x, and increase/decrease in expression between 4x and 8x; (4) Increase/decrease in expression between 2x, 4x and 8x – i.e. linear ; (5) Increase/decrease in expression between 2x and 4x, and decrease/increase in expression between 4x and 8x – i.e. switch in direction of change. Expression was considered to increase/decrease if the absolute log_2(fold-change) between expression values for the ploidy pair was >1 (i.e. there was a >2-fold difference in expression).

3.2.4.3.8 Correlation analysis

Correlation between gene expression levels and ploidy, Spearman's ranked correlation coefficients (ρ) and significance values (p) were estimated for each gene for both genotypes separately using the 'cor.test' function in R. To avoid unequal sample numbers across the three ploidy levels, for diploids, data from $2x_{\text{-untreated}}$ was excluded for the correlation tests (i.e. only normalized gene counts for $Ler_{\text{-colchicine}}$ and $Sf_{\text{-colchicine}}$ were used). Since individual correlation tests were performed for each gene, to circumvent obtaining spurious correlations by random chance due to multiple comparisons, a cut-off was calculated for each correlation test using a bootstrapping approach (Efron, 1979; Efron & Hastie, 2016; Cochran, 2019). Briefly, a Spearman correlation test for expression v/s ploidy was performed 10,000 times for each gene using randomly resampled expression values for the gene each time to obtain a bootstrap distribution of correlation coefficients and p-values. The 5^{th} percentile of the p-values (p_5), 5^{th} percentile of rho-values (ρ_5) and 95^{th} percentile of rho-values (ρ_{95}) were calculated from this distribution. Correlations and p-values from the original estimates were considered true if $\rho>\rho_{95}$ (true positive correlations) or $\rho<\rho_5$ (true negative correlations), and $p< p_5$ (true p-values). Genes that passed the thresholds were declared as significantly correlated genes (SCGs).

3.2.4.3.9 Differential gene expression analysis

To identify genes that were differentially expressed across genotype and ploidies, we used the R package EdgeR (Robinson *et al.*, 2010; McCarthy *et al.*, 2012). The normalized count data for all samples was analysed using the following multivariate linear model matrix: *Genotype + Treatment + Ploidy + Genotype:Ploidy* created using the 'model.matrix' function in EdgeR. The variance of the data (biological, technical and gene-wise) was estimated using the function 'estimateDisp' that uses a Cox-Reid profile-

adjusted likelihood to estimate dispersions in data with multiple factors (McCarthy *et al.*, 2012). The normalized counts, dispersion estimates and design matrix were passed through the 'glmFit' function in EdgeR that fits a negative binomial generalized log-linear model to the read counts for each gene, the output of which is passed through the 'glmLRT' function in edgeR which uses an empirical Bayes shrinkage approach to estimate the dispersion parameter and a likelihood ratio test to obtain p-values. Since we conducted a large number of comparisons, we declared a gene as differentially expressed (DEGs) when the False Discovery Rate (FDR) value (p-value corrected for multiple comparisons) was <0.05.

A second EdgeR analysis model was run, using a subset of the data separated by genotype. Normalized gene counts for Ler samples and Sf samples were used to determine differentially expressed genes for the two genotypes individually to compare gene expression differences between genotypes in response to increased ploidy using an additive linear model: *Treatment + Ploidy* following the same method as described above.

While the former analysis model gives a list of DEGs that change between the genotypes, treatments, ploidies and genotype-by-ploidy interaction, the latter provides insight to gene expression changes within the specific genotypes in response to increased ploidy. By drawing comparisons between the results obtained from the first analysis and those obtained from the second, genes that respond to ploidy irrespective of the genotype and those that respond to ploidy depending on their genetic background can be identified.

3.2.4.3.10 Functional characterisation

Functional characterisation of differentially expressed genes (DEGs) and significantly correlated genes (SCGs) was done using Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. DEG and SCG lists for each analysis model were independently parsed through the functional annotation tool in Database for Annotation, Visualization and Integrated Discovery (DAVID v6.8; https://david.ncifcrf.gov) for GO enrichment and KEGG pathway analyses. GO terms and KEGG pathways were considered significantly enriched if the adjusted p-value (Benjamini-Hochberg adjusted for multiple corrections) was <0.05.

A list of genes functionally associated with the phenotyped traits (described previously) were downloaded from TAIR (https://www.arabidopsis.org; accessed 01/07/2019) using the keywords : (1) Trichome differentiation, trichome morphogenesis and trichome branching – for trichome trait; (2) Guard cell differentiation, stomatal complex morphogenesis, stomatal complex development – for stomatal trait; (3) Photoperiodism, flowering time, regulation of photoperiodism – for flowering time trait; (4) Leaf morphogenesis, leaf shaping, leaf formation, leaf vascular tissue pattern formation, leaf development, regulation of leaf development – for leaf traits. These lists were then scanned for overlaps with the DEG and SCG lists to help elucidate if the DEGs or SCGs obtained explained the phenotypic changes observed. Genes associated with other phenotypes like seed size and number, petal area, and pollen area were not studied as our gene expression data is derived from leaf tissues which are unlikely to have a direct association with these traits.

3.3 RESULTS

3.3.1 The effects of increasing ploidy were significant and along the same direction for most phenotypic traits

Summary statistics (mean, standard deviation and coefficient of variance) for all phenotypic traits studied for all lines are presented in Table 3.1. Figure 3.2 summarizes the effect of three ploidy levels (diploids, tetraploids and octoploids) across two genetic backgrounds for various phenotypic traits.

The two-way ANOVA on untreated diploids ($2x_{\text{-untreated}}$), colchicine treated diploids ($2x_{\text{-colchicine}}$), colchicine treated tetraploids ($4x_{\text{-colchicine}}$) and colchicine treated octoploids ($8x_{\text{-colchicine}}$) testing for the effects of colchicine treatment, genotype, ploidy and genotype-by-ploidy interaction on various phenotypic traits showed that ploidy had a significant effect on all traits (Table 3.2) and significant genotype-by-ploidy interaction for four traits. A Tukey's HSD post-hoc test was conducted for comparisons that had significant ANOVAs, and results of the output are presented in Table 3.3.

Table 3.1 | Summary statistics for nine phenotypic traits across three ploidy levels and two genotypes. Means, standard deviations (SD), and coefficient of variances (CV) for two genotypes (Ler and Sf). 'u' corresponds to untreated plants and 'c' corresponds to colchicine treated plants. 2x, 4x and 8x refer to diploids, tetraploids and octoploids respectively.

Trait		Ler2x (u)	Ler2x (c)	Ler4x (c)	Ler8x (c)	Sf2x (u)	Sf2x (c)	Sf4x (c)	Sf8x (c)
Flowering Time (Days)	**Mean**	70.5	71.5	70.4	73.0	58.5	59.3	59.6	61.4
	SD	1.5	1.5	2.3	0.0	1.6	0.5	0.7	0.6
	CV	2.1	2.1	3.3	0.0	2.7	0.8	1.3	0.9
Leaf Area (mm^2)	**Mean**	630.1	491.9	555.6	327.9	454.0	551.9	463.2	219.8
	SD	97.4	75.5	59.0	86.3	88.4	43.6	88.8	55.4
	CV	15.5	15.3	10.6	26.3	19.5	7.9	19.2	25.2
Leaf Ratio (Length: Width)	**Mean**	2.0	1.8	1.6	1.3	2.5	2.4	1.9	1.9
	SD	0.1	0.1	0.1	0.1	0.2	0.2	0.3	0.4
	CV	2.8	3.4	7.7	4.7	6.7	7.7	13.5	19.4
Average Petal Area (mm^2)	**Mean**	2.0	1.7	3.0	3.7	1.6	1.7	3.1	4.2
	SD	0.4	0.4	0.3	0.9	0.2	0.2	0.2	0.2
	CV	19.9	22.6	10.5	23.2	14.4	12.3	6.2	4.2
Average Pollen Area (μm^2)	**Mean**	552.9	610.9	868.1	1240.6	515.2	531.2	878.9	1116.1
	SD	24.3	141.8	58.1	70.5	66.7	59.0	87.8	257.0
	CV	4.4	23.2	6.7	5.7	13.0	11.1	10.0	23.0
Rosette Diameter (cm)	**Mean**	108.3	113.3	96.8	53.0	130.1	130.9	126.0	87.4
	SD	9.3	7.4	6.1	7.3	4.1	5.1	10.2	14.8
	CV	8.6	6.5	6.3	13.7	3.1	3.9	8.1	16.9
Average Seed Number (per fruit)	**Mean**	62.2	68.8	23.2	3.7	72.9	72.5	47.2	7.5
	SD	8.5	6.7	7.5	1.5	4.4	10.8	9.7	2.1
	CV	13.6	9.7	32.3	40.2	6.1	15.0	20.5	27.6
Average Seed Size (mm^2)	**Mean**	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.2
	SD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	CV	9.3	4.9	6.4	11.3	4.7	4.4	6.5	21.2
Stomatal Diameter (µm)	**Mean**	22.7	26.1	30.4	36.6	21.1	22.2	32.4	41.5
	SD	1.5	1.4	1.8	4.1	1.5	1.6	1.7	3.7
	CV	6.4	5.2	5.9	11.1	6.9	7.3	5.3	8.9

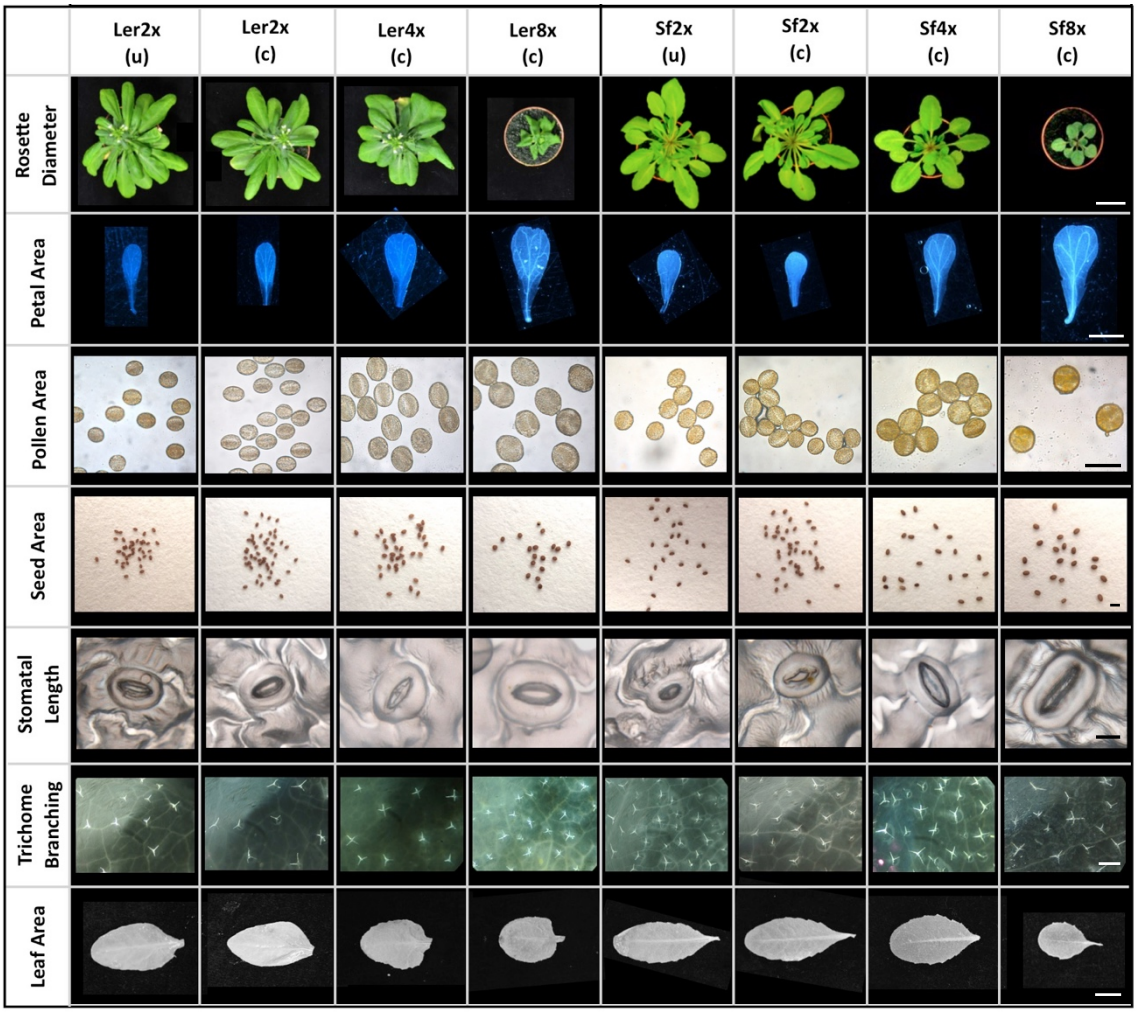


Figure 3.2 | Summary of phenotypes observed for various phenotypic traits across three ploidy levels and for two genotypes. 'u' corresponds to untreated plants and 'c' corresponds to colchicine treated plants. 2x, 4x and 8x refer to diploids, tetraploids and octoploids respectively. Scale bars – Rosette Diameter (25mm); Petal Area (1mm); Pollen Area (50µm); Seed Area (1mm); Stomatal Length (10µm); Trichome Branching (1mm); Leaf Area (10mm).

Table 3.2 | ANOVA output for each phenotype. Two-way ANOVA results for the effect of genotype, ploidy, genotype-by-ploidy interaction, and colchicine treatment within ploidy on various phenotypic traits. Treatment is nested within ploidy (Ploidy/Treatment) as only one ploidy level (diploid) has two levels of treatment (untreated and colchicine-treated). Genotype:Ploidy represents the genotype-by-ploidy interaction term. Anova model = *Trait ~ Genotype + Ploidy + (Ploidy/Colchicine) + (Genotype x Ploidy)*

Trait	Factor	Df	Sum Sq	F value	Pr(>F)
	Genotype	1	1969.72	1062.18	**8.155E-36**

Flowering Time (Days)	Ploidy	2	41.63	11.22	**9.131E-05**
	Ploidy/Treatment	1	6.12	3.30	7.503E-02
	Genotype:Ploidy	2	5.05	1.36	2.653E-01
Leaf Area (mm^2)	Genotype	1	62646.63	8.05	**7.718E-03**
	Ploidy	2	474338.43	30.48	**3.177E-08**
	Ploidy/Treatment	1	2039.64	0.26	6.121E-01
	Genotype:Ploidy	2	4751.50	0.31	7.390E-01
Leaf Ratio (Leaf Length:Leaf Width)	Genotype	1	2.23	65.17	**2.569E-09**
	Ploidy	2	2.60	38.00	**2.742E-09**
	Ploidy/Treatment	1	0.12	3.57	6.778E-02
	Genotype:Ploidy	2	0.09	1.36	2.712E-01
Average Petal Area (mm^2)	Genotype	1	0.22	1.72	1.964E-01
	Ploidy	2	37.84	150.84	**1.250E-19**
	Ploidy/Treatment	1	0.07	0.57	4.553E-01
	Genotype:Ploidy	2	1.02	4.05	**2.486E-02**
Average Pollen Area (μm^2)	Genotype	1	37519.30	3.38	7.166E-02
	Ploidy	2	3326708.81	150.02	**4.331E-22**
	Ploidy/Treatment	1	10951.26	0.99	3.250E-01
	Genotype:Ploidy	2	29249.15	1.32	2.764E-01
Rosette Diameter (mm)	Genotype	1	9241.75	138.65	**4.980E-16**
	Ploidy	2	19217.82	144.15	**1.743E-21**
	Ploidy/Treatment	1	54.13	0.81	3.718E-01
	Genotype:Ploidy	2	501.77	3.76	**3.002E-02**
Average Seed Number (per fruit)	Genotype	1	30.91	0.54	4.707E-01
	Ploidy	2	23034.83	199.43	**6.734E-17**
	Ploidy/Treatment	1	33.49	0.58	4.530E-01
	Genotype:Ploidy	2	476.67	4.13	**2.730E-02**
Average Seed Size (mm^2)	Genotype	1	0.02	59.36	**2.754E-08**
	Ploidy	2	0.03	43.76	**3.377E-09**
	Ploidy/Treatment	1	0.00	0.03	8.700E-01
	Genotype:Ploidy	2	0.00	0.80	4.587E-01
Stomatal Diameter (µm)	Genotype	1	9.50	1.62	2.085E-01
	Ploidy	2	2550.75	217.87	**2.062E-25**
	Ploidy/Treatment	1	53.34	9.11	**3.989E-03**
	Genotype:Ploidy	2	143.82	12.28	**4.575E-05**

Table 3.3 | Output table of Tukey's HSD test for ANOVAs that tested significant for the effects of ploidy and genotype-by-ploidy nine phenotypic traits. Lower and upper bound represent the lower and upper limits of the 95% confidence interval. Sig represents the adjusted P-value for the comparison corrected for multiple comparisons.

Trait	Factor	Comparison	Mean Diff	Lower Bound	Upper Bound	Sig
Flowering Time (Days)	Ploidy	4-2	-0.52	-1.53	0.49	4.30E-01
	Ploidy	8-2	1.68	0.49	2.87	3.67E-03
	Ploidy	8-4	2.20	0.87	3.53	5.78E-04
Leaf Area (mm2)	Ploidy	4-2	30.92	-52.91	114.76	6.41E-01
	Ploidy	8-2	-204.60	-288.43	-120.76	2.94E-06
	Ploidy	8-4	-235.52	-332.32	-138.72	3.10E-06
Leaf Ratio (Length: Width)	Ploidy	4-2	-0.21	-0.39	-0.04	1.57E-02
	Ploidy	8-2	-0.35	-0.53	-0.18	7.03E-05
	Ploidy	8-4	-0.14	-0.34	0.06	2.21E-01
Average Petal Area (mm^2)	Ploidy	4-2	0.61	0.31	0.90	2.89E-05
	Ploidy	8-2	1.51	1.18	1.84	1.07E-12
	Ploidy	8-4	0.90	0.53	1.28	1.92E-06
	G x P	Sf x 2-Ler x 2	-0.65	-1.08	-0.23	5.96E-04
	G x P	Ler x 4-Ler x 2	0.19	-0.39	0.77	9.22E-01
	G x P	Sf x 4-Ler x 2	0.26	-0.24	0.76	6.32E-01
	G x P	Ler x 8-Ler x 2	0.90	0.28	1.53	1.33E-03
	G x P	Sf x 8-Ler x 2	1.36	0.78	1.94	2.48E-07
	G x P	Ler x 4-Sf x 2	0.84	0.30	1.38	4.75E-04
	G x P	Sf x 4-Sf x 2	0.91	0.46	1.37	7.13E-06
	G x P	Ler x 8-Sf x 2	1.55	0.96	2.15	1.56E-08
	G x P	Sf x 8-Sf x 2	2.01	1.47	2.55	1.87E-12
	G x P	Sf x 4-Ler x 4	0.07	-0.53	0.67	9.99E-01
	G x P	Ler x 8-Ler x 4	0.71	0.00	1.42	4.90E-02
	G x P	Sf x 8-Ler x 4	1.17	0.50	1.84	8.07E-05
	G x P	Ler x 8-Sf x 4	0.64	-0.01	1.29	5.41E-02
	G x P	Sf x 8-Sf x 4	1.09	0.49	1.70	4.02E-05
	G x P	Sf x 8-Ler x 8	0.45	-0.26	1.16	4.11E-01
Average Pollen Area (μm^2)	Ploidy	4-2	172.57	94.74	250.39	6.15E-06
	Ploidy	8-2	477.41	385.32	569.50	0.00E+00
	Ploidy	8-4	304.84	202.38	407.31	8.39E-09
Rosette Diameter (mm)	Ploidy	4-2	-1.84	-7.91	4.23	7.46E-01
	Ploidy	8-2	-43.01	-50.19	-35.84	0.00E+00
	Ploidy	8-4	-41.18	-49.12	-33.23	0.00E+00
	G x P	Sf x 2-Ler x 2	20.36	11.66	29.05	1.11E-07
	G x P	Ler x 4-Ler x 2	-6.53	-17.12	4.06	4.58E-01
	G x P	Sf x 4-Ler x 2	22.72	12.13	33.31	8.91E-07
	G x P	Ler x 8-Ler x 2	-50.28	-62.77	-37.79	0.00E+00
	G x P	Sf x 8-Ler x 2	-15.88	-28.37	-3.39	5.53E-03
	G x P	Ler x 4-Sf x 2	-26.89	-37.36	-16.41	1.01E-08
	G x P	Sf x 4-Sf x 2	2.36	-8.11	12.84	9.85E-01

	G x P	Ler x 8-Sf x 2	-70.64	-83.03	-58.24	0.00E+00
	G x P	Sf x 8-Sf x 2	-36.24	-48.63	-23.84	2.33E-10
	G x P	Sf x 4-Ler x 4	29.25	17.15	41.35	4.88E-08
	G x P	Ler x 8-Ler x 4	-43.75	-57.54	-29.96	1.40E-11
	G x P	Sf x 8-Ler x 4	-9.35	-23.14	4.44	3.52E-01
	G x P	Ler x 8-Sf x 4	-73.00	-86.79	-59.21	0.00E+00
	G x P	Sf x 8-Sf x 4	-38.60	-52.39	-24.81	8.68E-10
	G x P	Sf x 8-Ler x 8	34.40	19.10	49.70	2.98E-07
Average Seed Number (per fruit)	Ploidy	4-2	-17.39	-25.40	-9.38	3.15E-05
	Ploidy	8-2	-49.51	-57.51	-41.50	2.66E-14
	Ploidy	8-4	-32.12	-41.54	-22.70	1.36E-08
	G x P	Sf x 2-Ler x 2	13.67	2.62	24.71	8.98E-03
	G x P	Ler x 4-Ler x 2	-25.94	-41.27	-10.62	2.45E-04
	G x P	Sf x 4-Ler x 2	-1.95	-14.70	10.80	9.97E-01
	G x P	Ler x 8-Ler x 2	-45.43	-60.76	-30.10	1.54E-08
	G x P	Sf x 8-Ler x 2	-41.64	-54.40	-28.89	2.02E-09
	G x P	Ler x 4-Sf x 2	-39.61	-55.37	-23.84	3.96E-07
	G x P	Sf x 4-Sf x 2	-15.62	-28.89	-2.34	1.42E-02
	G x P	Ler x 8-Sf x 2	-59.10	-74.86	-43.33	9.72E-11
	G x P	Sf x 8-Sf x 2	-55.31	-68.58	-42.04	8.69E-12
	G x P	Sf x 4-Ler x 4	23.99	6.99	40.99	2.33E-03
	G x P	Ler x 8-Ler x 4	-19.49	-38.50	-0.48	4.21E-02
	G x P	Sf x 8-Ler x 4	-15.70	-32.71	1.30	8.29E-02
	G x P	Ler x 8-Sf x 4	-43.48	-60.48	-26.48	2.85E-07
	G x P	Sf x 8-Sf x 4	-39.69	-54.42	-24.97	1.03E-07
	G x P	Sf x 8-Ler x 8	3.79	-13.22	20.79	9.82E-01
Average Seed Size (mm^2)	Ploidy	4-2	0.03	0.01	0.05	4.43E-04
	Ploidy	8-2	0.05	0.03	0.06	5.66E-06
	Ploidy	8-4	0.01	-0.01	0.03	3.58E-01
Stomatal Diameter (µm)	Ploidy	4-2	2.76	0.81	4.71	3.53E-03
	Ploidy	8-2	10.15	8.29	12.01	0.00E+00
	Ploidy	8-4	7.39	5.18	9.61	3.95E-10
	G x P	Sf x 2-Ler x 2	-1.69	-4.36	0.99	4.34E-01
	G x P	Ler x 4-Ler x 2	0.93	-2.75	4.60	9.75E-01
	G x P	Sf x 4-Ler x 2	2.94	-0.16	6.04	7.29E-02
	G x P	Ler x 8-Ler x 2	7.17	4.07	10.28	1.54E-07
	G x P	Sf x 8-Ler x 2	11.99	8.74	15.24	0.00E+00
	G x P	Ler x 4-Sf x 2	2.61	-1.16	6.38	3.29E-01
	G x P	Sf x 4-Sf x 2	4.63	1.41	7.85	1.22E-03
	G x P	Ler x 8-Sf x 2	8.86	5.64	12.08	1.46E-09
	G x P	Sf x 8-Sf x 2	13.68	10.32	17.04	0.00E+00
	G x P	Sf x 4-Ler x 4	2.02	-2.07	6.10	6.90E-01

G x P	Ler x 8-Ler x 4	6.25	2.16	10.33	5.06E-04
G x P	Sf x 8-Ler x 4	11.07	6.87	15.26	4.84E-09
G x P	Ler x 8-Sf x 4	4.23	0.65	7.82	1.21E-02
G x P	Sf x 8-Sf x 4	9.05	5.34	12.76	3.91E-08
G x P	Sf x 8-Ler x 8	4.82	1.11	8.53	4.34E-03

A two-way ANOVA on flowering time data showed that genotype and ploidy had significant effects, and that there was no genotype-by-ploidy interaction. The differences between untreated and colchicine treated diploids were not significant (Table 3.2). A Tukey's post hoc test conducted on significant ANOVA factors showed that there was significant difference between tetraploids and octoploids, and between diploids and octoploids, however, the differences between diploids and tetraploids was not significant (Table 3.3). Octoploids flowered significantly later than their respective diploids and tetraploids, and plants belonging to Ler flowered later than those belonging to Sf (Figure 3.3 and Table 3.1).

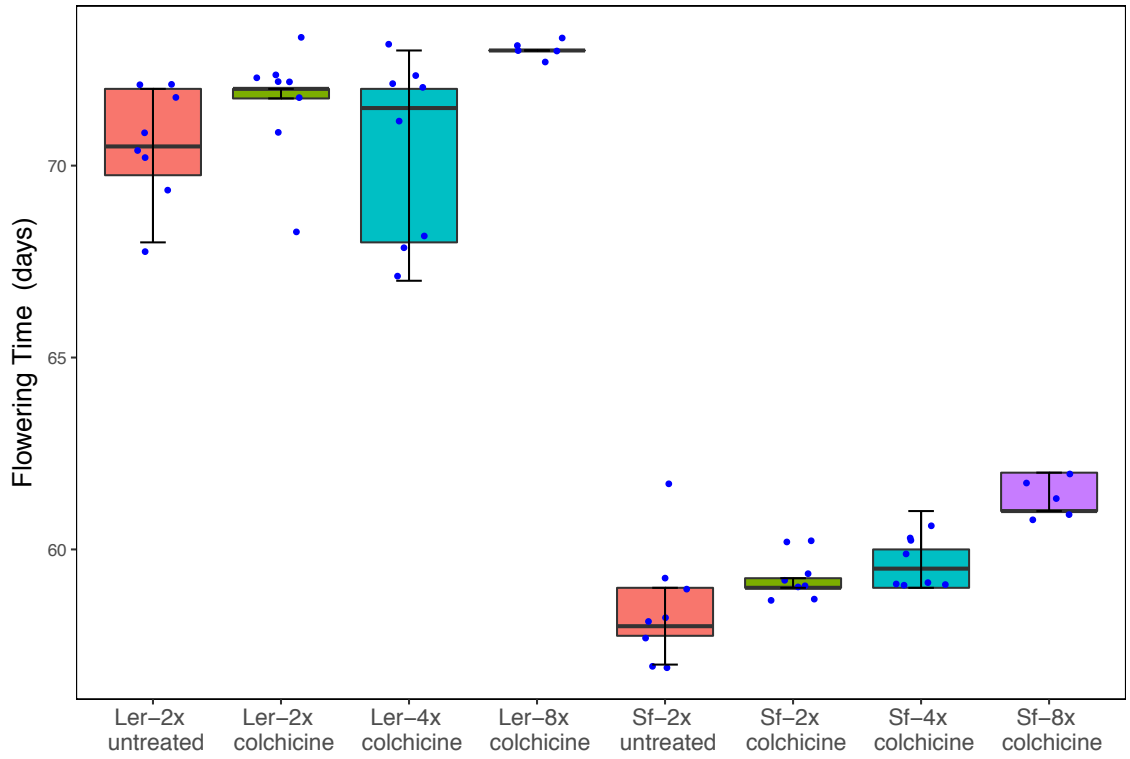


Figure 3.3 | Flowering time data for diploids, tetraploids and octoploids of Ler and Sf. Flowering time was measured in days taken for the appearance of the first white petals post germination. Boxes represent the extremities of the interquartile range. The horizontal black line within the boxes represents the median flowering time for each line, whiskers indicate the minimum and maximum values and blue dots represent actual data points for each plant within the line. N=8 replicates per line.

A two-way ANOVA on rosette diameter data showed that genotype, ploidy and genotype-by-ploidy interaction had significant effects. The differences between untreated and colchicine treated diploids were not significant (Table 3.2). A Tukey's post hoc test conducted on significant ANOVA factors showed that there was significant difference between tetraploids and octoploids, and between diploids and octoploids, however, the differences between diploids and tetraploids was not significant (Table 3.3). Octoploids had smaller rosettes in comparison to diploids and tetraploids for both genotypes (Figure 3.4 and Table 3.1). However, the magnitude of difference between tetraploids and octoploids was greater for Ler in comparison to Sf, illustrating the significant genotype-by-ploidy effects found in the ANOVA data.

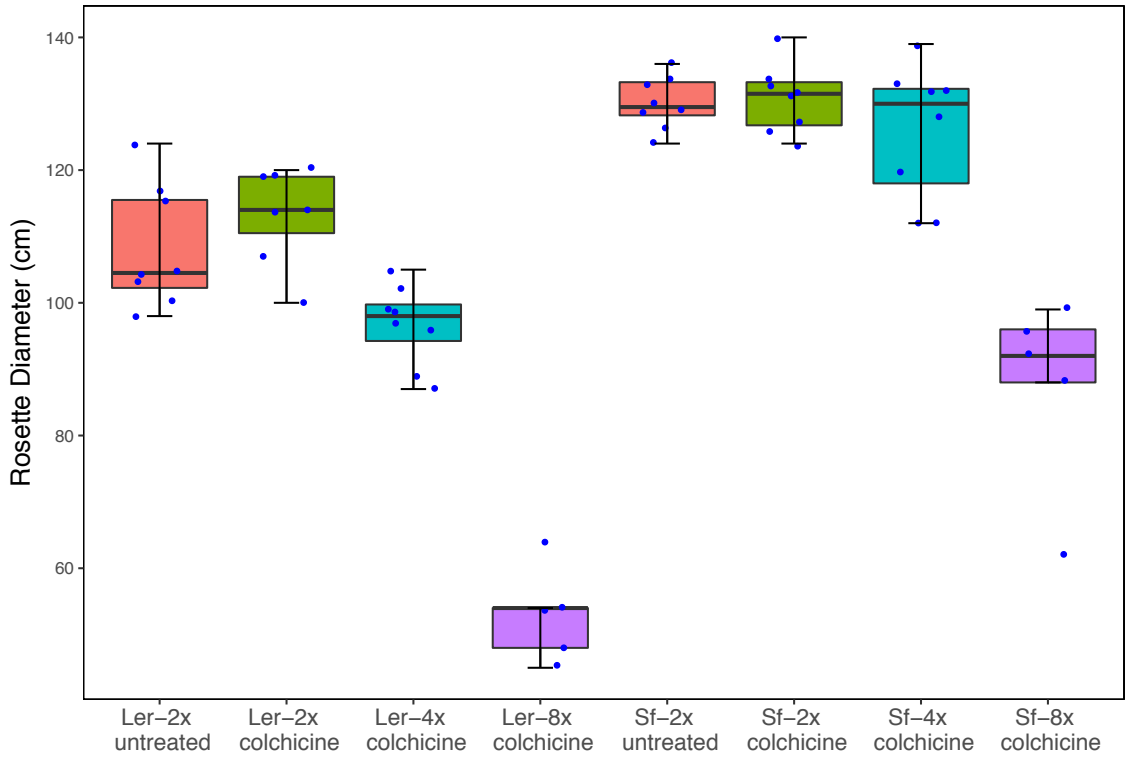


Figure 3.4 | Rosette diameter data for diploids, tetraploids and octoploids of Ler and Sf. Rosette diameter was measured along the longest axis. Boxes represent the extremities of the interquartile range. The horizontal black line within the boxes represents the median rosette diameter for each line, whiskers indicate the minimum and maximum values and blue dots represent actual data points for each plant within the line. N=8 replicates per line.

While the effect of ploidy flowering time and rosette diameter appears to be enhanced between tetraploids and octoploids, for petal area, the effect was larger between diploids and tetraploids (Figure 3.5, Table 3.1). Petal area increased with increasing ploidy. A two-way ANOVA on petal area data showed that genotype, ploidy and genotype-by-ploidy interaction had significant effects. The differences between untreated and colchicine treated diploids were not significant (Table 3.2). A Tukey's post hoc test conducted on significant ANOVA factors showed that there were significant difference between all ploidy comparisons (Table 3.3).

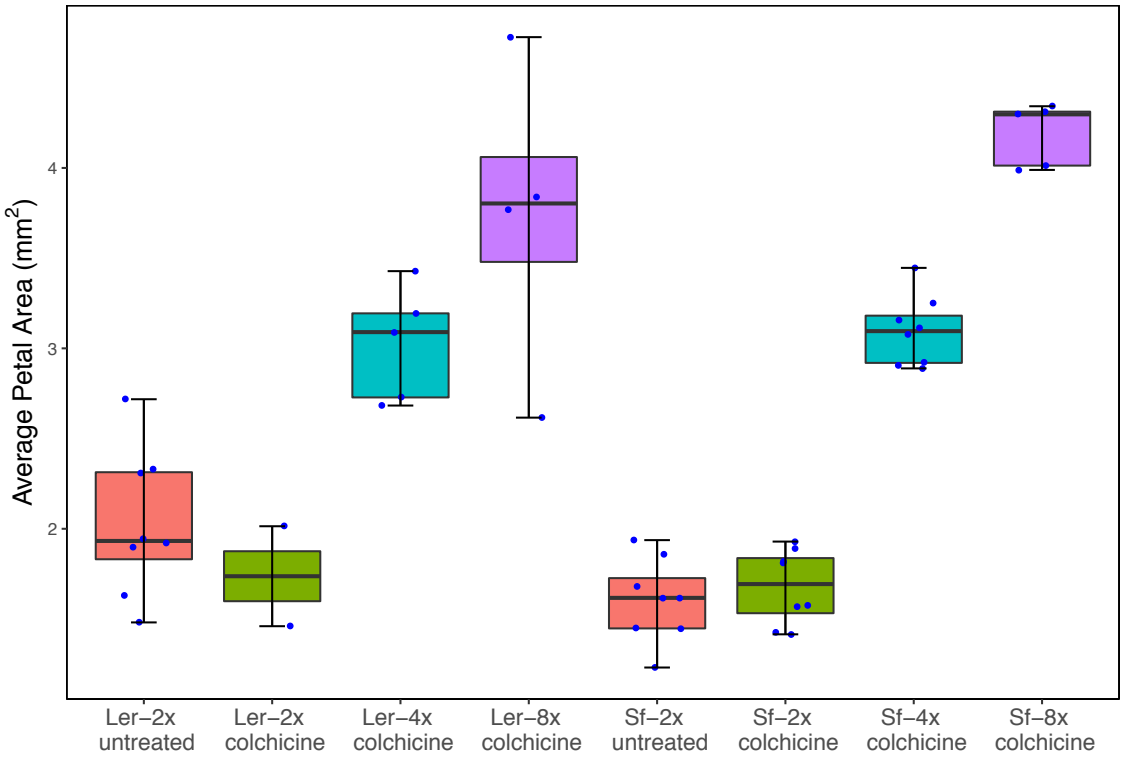


Figure 3.5 | Average petal area data for diploids, tetraploids and octoploids of Ler and Sf. Each data point represents the averaged petal area across the four petals from each flower. Boxes represent the extremities of the interquartile range. The horizontal black line within the boxes represents the median petal area for each line, whiskers indicate the minimum and maximum values and blue dots represent actual data points for each plant within the line. N=8 replicates per line.

A two-way ANOVA on average pollen area data showed that genotype and ploidy had significant effects, and that there was no genotype-by-ploidy interaction. The differences between untreated and colchicine treated diploids were not significant (Table 3.2). A Tukey's post hoc test conducted on significant ANOVA factors showed that there were significant differences between all ploidy comparisons (Table 3.3). Pollen area increases with increased ploidy (Figure 3.6).

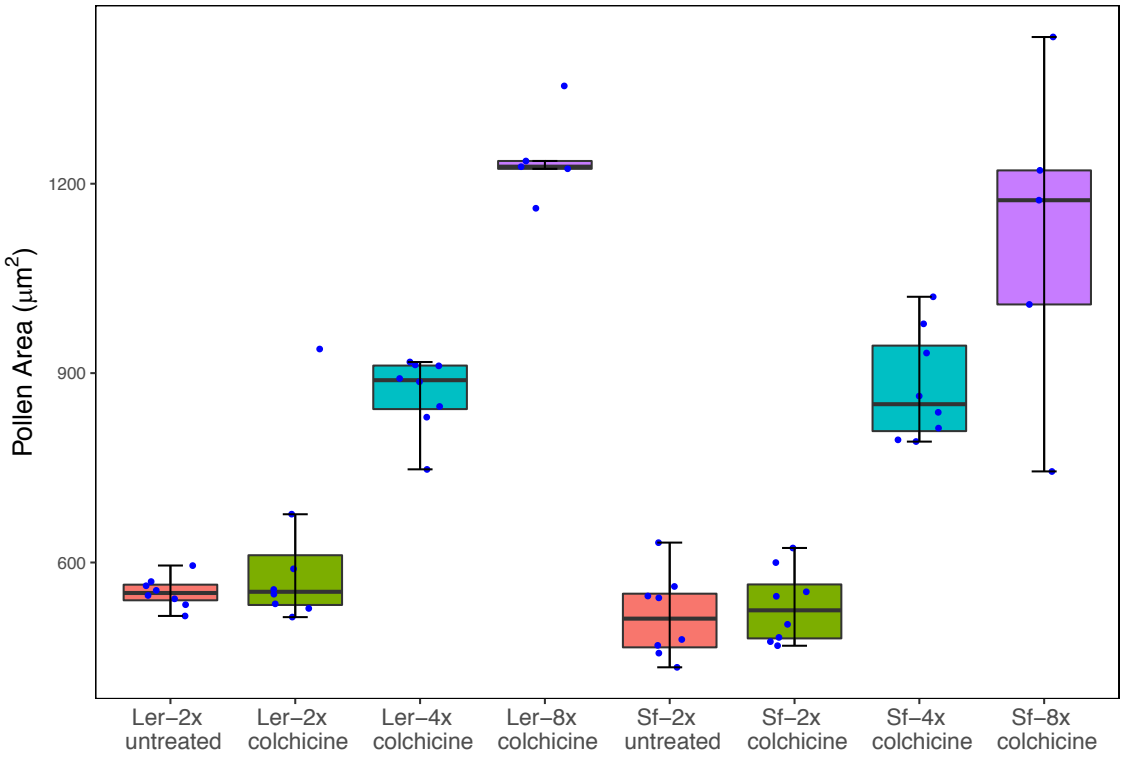


Figure 3.6 | Average pollen area data for diploids, tetraploids and octoploids of Ler and Sf. Each data point represents the averaged pollen area across 10-30 hydrated pollen from each flower. Boxes represent the extremities of the interquartile range. The horizontal black line within the boxes represents the median pollen area for each line, whiskers indicate the minimum and maximum values and blue dots represent actual data points for each plant within the line. N=8 replicates per line.

Seed size increased with increasing ploidy (Table 3.1, Figure 3.7). The two-way ANOVA data on seed size showed that genotype and ploidy had effects on seed size but there was no genotype-by-ploidy interaction (Table 3.2). A Tukey's post hoc test conducted on significant ANOVA factors showed that there was significant difference between diploids and tetraploids and diploids and octoploids, however the difference between tetraploids and octoploids was not significant (Table 3.3). Octoploid seeds, despite having similar surface area to tetraploids, could be distinguished based on their morphology as the seeds were more spherical in comparison to diploid seeds that were more elliptical (Figure 3.2).

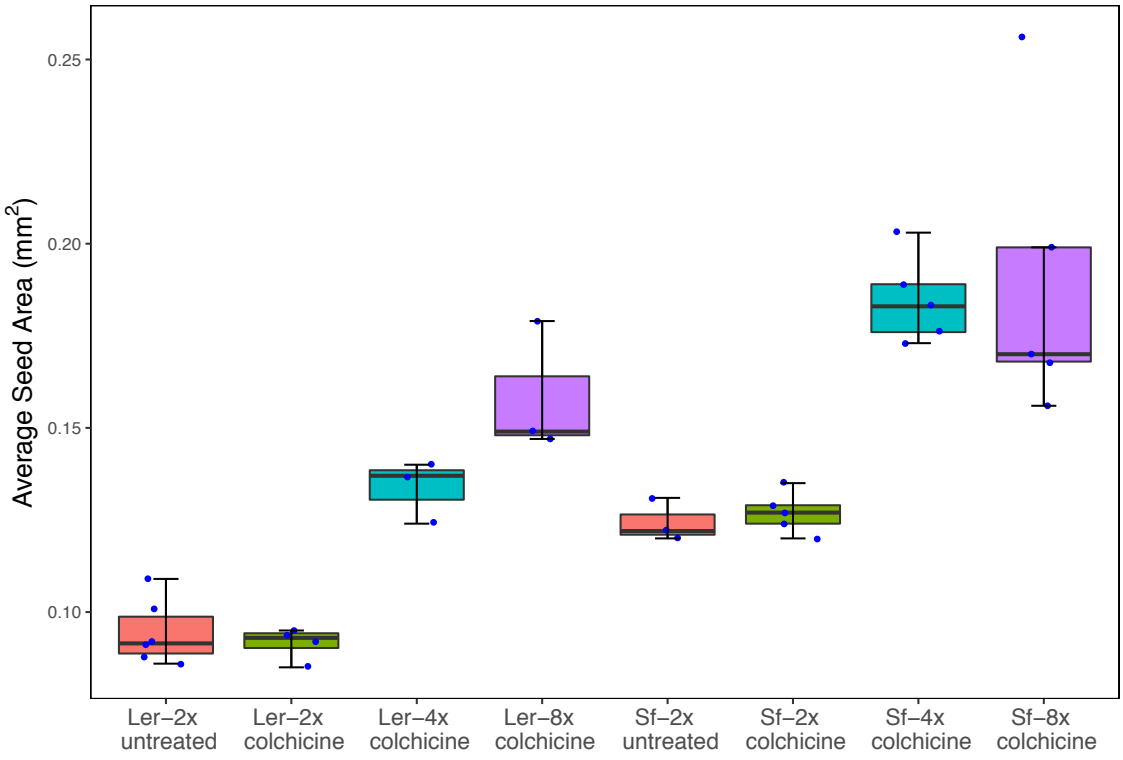


Figure 3.7 | Average seed area data for diploids, tetraploids and octoploids of Ler and Sf. Each data point represents the averaged seed area across all seeds from five siliques (fruits) from each plant. Each silique had between 5-80 seeds depending on the ploidy. Boxes represent the extremities of the interquartile range. The horizontal black line within the boxes represents the median seed surface area for each line, whiskers indicate the minimum and maximum values and blue dots represent actual data points for each plant within the line. N=5 data points per line.

While seed area increased with increasing ploidy, the seed number (i.e. the average number of seeds in each silique) significantly decreased (Table 3.1, Figure 3.8). The magnitude of differences between the ploidies varied between Ler and Sf, depicting the significant ploidy and genotype-by-ploidy interaction effects found in the ANOVA (Table 3.2). A Tukey's post hoc test conducted on significant ANOVA factors showed that all ploidy comparisons were significant (Table 3.3).

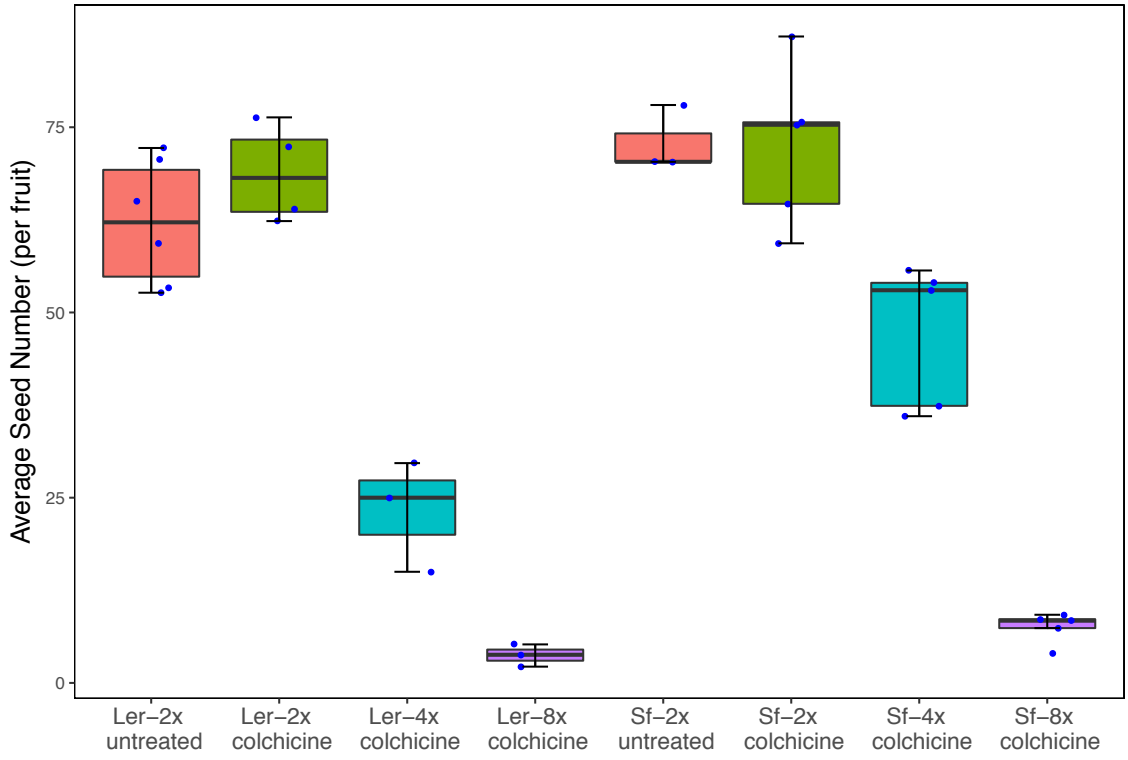


Figure 3.8 | Average seed number data for diploids, tetraploids and octoploids of Ler and Sf. Each data point represents the average number of seeds per fruits calculated from five siliques (fruits) from each plant. Boxes represent the extremities of the interquartile range. The horizontal black line within the boxes represents the median seed number for each line, whiskers indicate the minimum and maximum values and blue dots represent actual data points for each plant within the line. N=5 data points per line.

The stomatal diameter increased with increasing ploidy (Table 3.1, Figure 3.9). Sf showed a larger increase in stomatal diameter in response to increased ploidy in comparison to Ler, thus showing a genotype-by-ploidy interaction effect. A two-way ANOVA on stomatal diameter data showed that colchicine, ploidy and genotype-by-ploidy interaction had significant effects. This was the only trait that showed no main effect of genotype and showed significant colchicine effects between the two diploids (Table 3.2). A Tukey's post hoc test conducted on significant ANOVA factors showed that all ploidy comparisons were significant for Sf but only tetraploid to octoploid, and diploid to octoploid for Ler (Table 3.3).

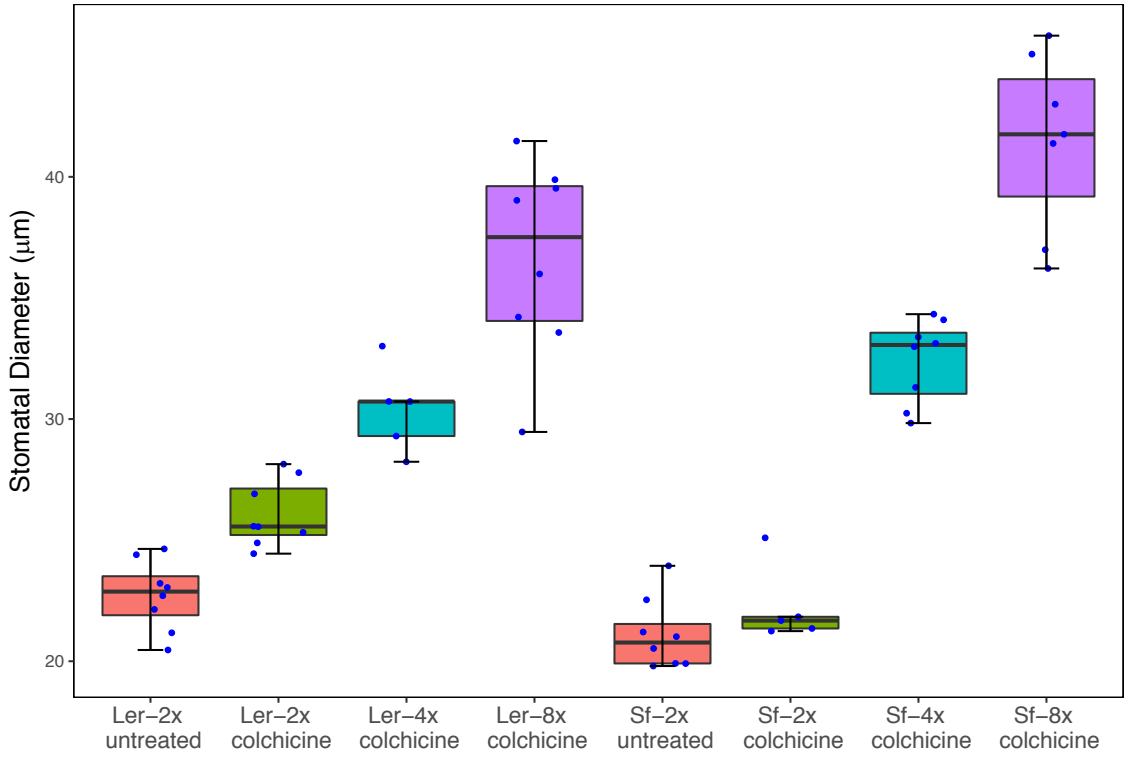
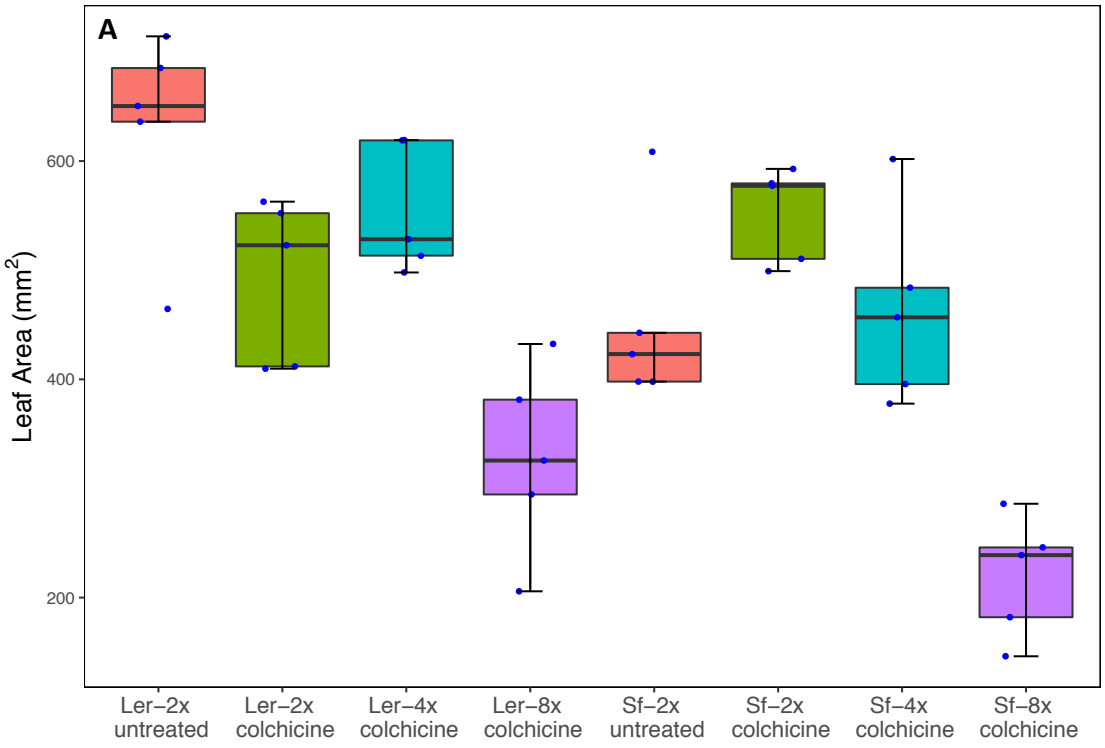


Figure 3.9 | Stomatal diameter data for diploids, tetraploids and octoploids of Ler and Sf. Each data point represents the averaged stomatal diameter across five stomata per cotyledon from each plant. Boxes represent the extremities of the interquartile range. The horizontal black line within the boxes represents the median stomatal diameter for each line, whiskers indicate the minimum and maximum values and blue dots represent actual data points for each plant within the line. N=8 data points per line.

Leaf area showed large variance (up to 26%) between replicates making it difficult to determine general patterns of response to ploidy from a limited number of replicates (Table 3.1, Figure 3.10A). The ANOVA data showed significant genotype and ploidy effects (Table 3.2). There was a significant decrease in leaf area between tetraploids and octoploids but not between diploids and tetraploids (Table 3.3). Morphologically, as with seed area, we found that the leaf shape changed with increasing ploidy, tending towards shorter and wider leaves (Figure 3.2). Thus, a better alternative to estimate the changes in leaf phenotype was to determine the leaf shape in the form of a ratio of leaf length to leaf width. Here, we saw a relatively clear pattern of decreasing leaf ratio with increasing ploidy (Table 3.1, Figure 3.10B).



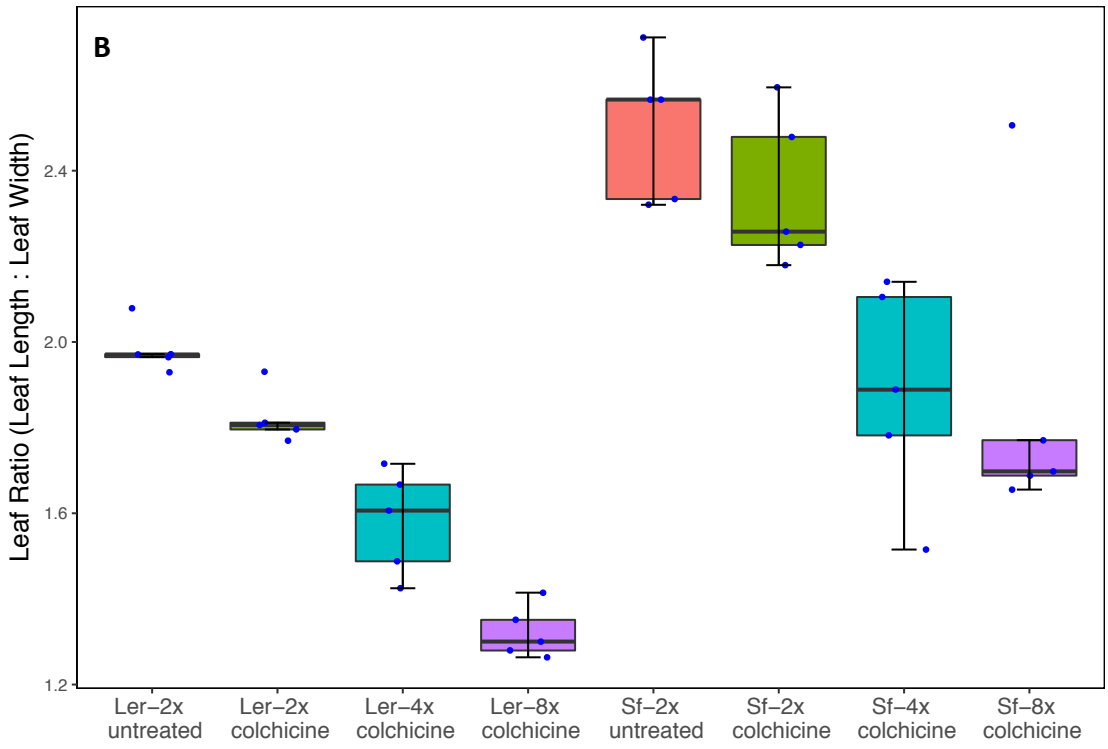


Figure 3.10 | Leaf morphology for diploids, tetraploids and octoploids of Ler and Sf. (A) Leaf area of the 5th rosette leaf was measured post flowering. (B) Leaf length to leaf width ratio was calculated from images of the 5th rosette leaf. Boxes represent the extremities of the interquartile range. The horizontal black line within the boxes represents the median values for each line, whiskers indicate the minimum and maximum values and blue dots represent actual data points for each plant within the line. N=5 data points per line.

Trichome branching, specifically, showed a distinct response to increased ploidy depending on the genotype (Figure 3.11). Trichomes of Ler diploids had a large proportion of 3 branches (untreated = 96.84%; treated = 94.79%) which decreases to 31.19% in tetraploids and 3.32% in octoploids, where 4 branches (67.35%) and 5 branches (80.12%) took predominance respectively. While Ler trichomes were characterised by an increase in the number of branches in octoploids, Sf exhibited a relatively subtle response in branching wherein only a small proportion of 5-branched trichomes (11.12%) are observed in octoploids. Thus, Ler primarily responded by a change in the number of trichome branches with increasing ploidy level whereas Sf largely responded by a change in proportion of 4-branched trichomes.

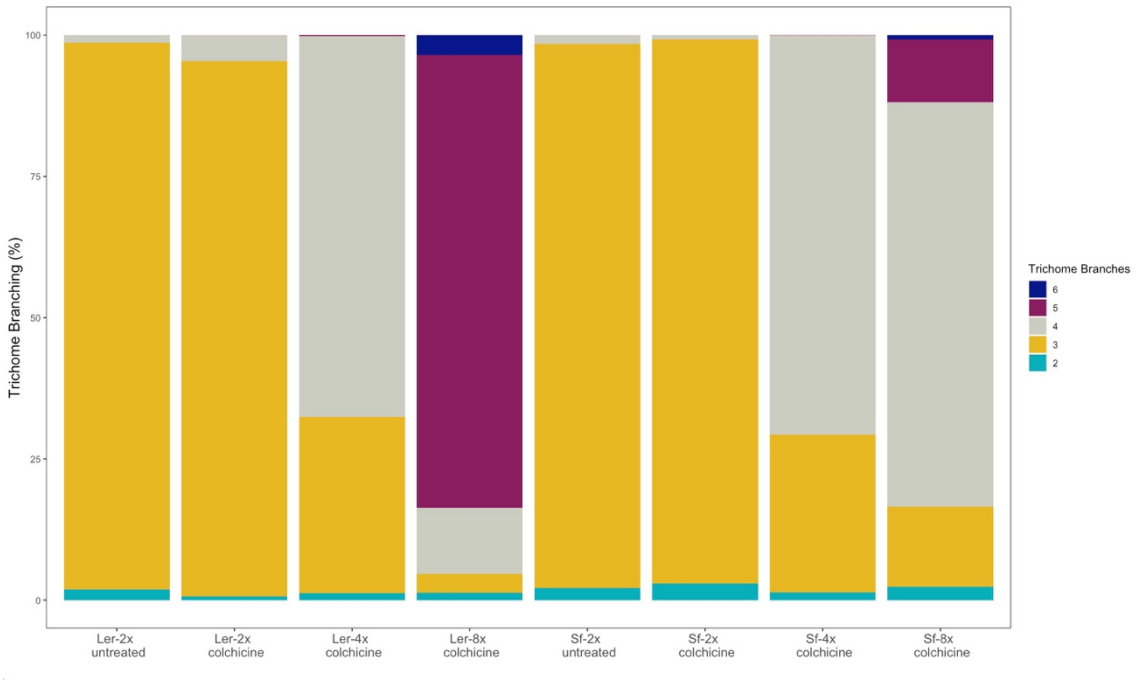


Figure 3.11 | Trichome branching data for diploids, tetraploids and octoploids of Ler and Sf. The total number of trichomes on the 5th rosette leaves were counted and categorised based on the number of branches the trichomes had. Trichome branching data is represented in each bar as a percentage of n-branched (n=2 to 6) trichomes of the total number of trichomes averaged across the replicates. N=5 replicates per line.

3.3.2 Exploratory analysis of gene expression data

3.3.2.1 Overview of read depth per sample

We obtained an average of 23.6 million paired end RNA-Seq reads per sample. After removal of low-quality reads using Trimmomatic (Bolger et al. 2014), an average of 21.3 million reads per sample remained (Table 3.4). Out of 27,655 genes in the reference annotation, our dataset had 3759 genes that were not expressed in Ler samples, 4067 genes that were not expressed in Sf samples and 2970 genes that were not expressed in any of the 24 samples.

Table 3.4 | Library information for the 24 sequenced samples. Three biological replicates were sequenced for each of the eight lines using Illumina Hiseq2500 to generate 126bp paired end reads. Illumina adapters, low quality sequences, and short sequences were removed using Trimmomatic v0.32 (Bolger et al. 2014). The number of read pairs obtained and retained post trimming are listed below for each sample.

Genotype	Ploidy	Group	No.of read pairs	Read pairs after trimming
Ler	2	Ler2x-untreated	23026032	19989316
Ler	2	Ler2x-colchicine	10340755	8250616

Ler	4	Ler4x-colchicine	23376099	20088505
Ler	8	Ler8x-colchicine	24483650	22106820
Sf	2	Sf2x-untreated	18025486	16187116
Sf	2	Sf2x-colchicine	27024911	24795384
Sf	4	Sf4x-colchicine	28718951	24668133
Sf	8	Sf8x-colchicine	20798788	18313736
Ler	2	Ler2x-untreated	29616263	27070446
Ler	2	Ler2x-colchicine	22668178	20367909
Ler	4	Ler4x-colchicine	31912627	29222196
Ler	8	Ler8x-colchicine	22049012	20533303
Sf	2	Sf2x-untreated	28413697	25879948
Sf	2	Sf2x-colchicine	23651740	18348521
Sf	4	Sf4x-colchicine	28338499	24760939
Sf	8	Sf8x-colchicine	26294372	23891891
Ler	2	Ler2x-untreated	23537183	22079952
Ler	2	Ler2x-colchicine	16649330	15684973
Ler	4	Ler4x-colchicine	22235371	21134113
Ler	8	Ler8x-colchicine	19594494	18425085
Sf	2	Sf2x-untreated	23422356	21958075
Sf	2	Sf2x-colchicine	24055638	22577082
Sf	4	Sf4x-colchicine	20663000	19505961
Sf	8	Sf8x-colchicine	26882548	25446721
		Total Reads	**565,778,980**	**511,286,741**
		Average Reads	**23,574,124**	**21,303,614**

3.3.2.2 Dimensionality reduction suggested that transcriptome profiles of all samples differed more between genotypes than between ploidy levels

Ordination analysis using a multidimensional scaling approach (MDS) on the biological coefficient of variance (BCV) as a metric showed clear separation of genotypes along component 1 (Figure 3.12 – x-axis) suggesting that transcriptional profiles of the samples differed between the genotypes. Along component 2 (Figure 3.12 – y-axis), there was good separation between diploid and octoploid samples for both genotypes, suggesting that octoploid transcriptome profiles were most dissimilar to diploid transcriptomes. Tetraploids of Ler were interspersed between diploids and octoploids suggesting that there is similarity in their transcriptomes to both diploids and octoploids. Sf tetraploids on the other hand showed similarity to diploid transcriptomes and there was complete separation between the tetraploid and octoploid Sf samples.

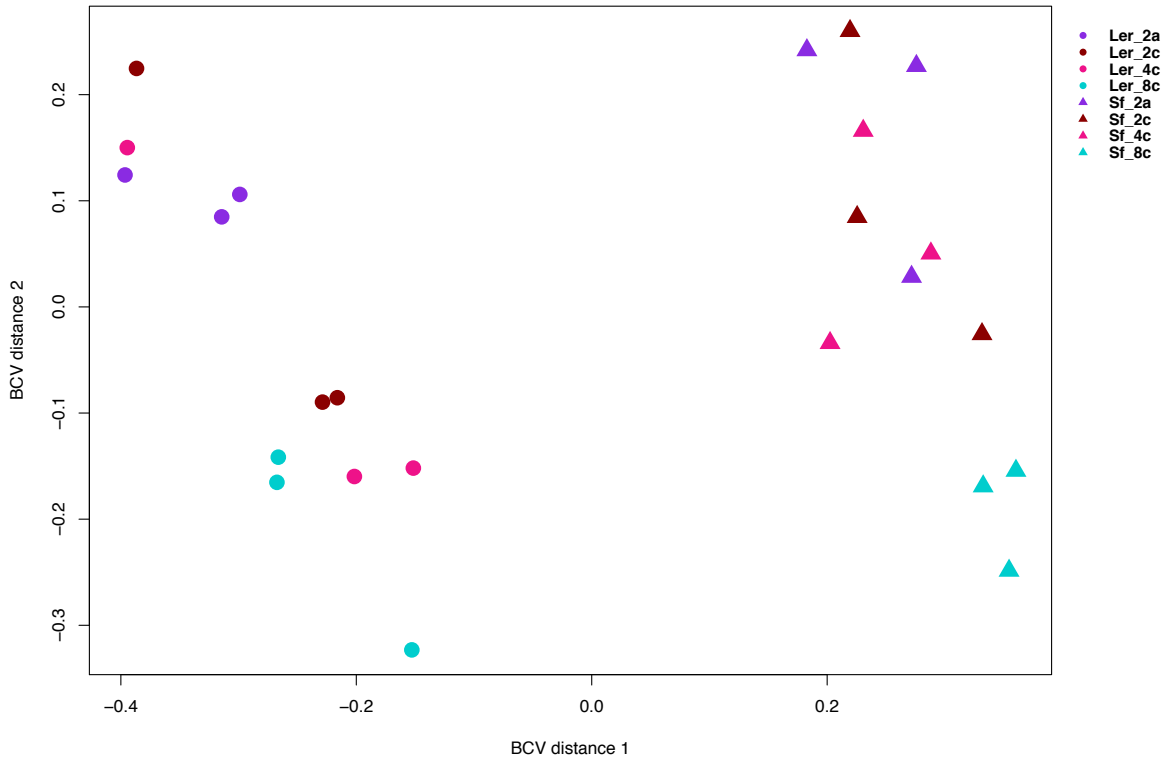


Figure 3.12 | Multidimensional Scaling (MDS) plot representing relationships between transcriptome profiles for the 24 samples. An ordination analysis was performed on normalized read counts by calculating distance measures based on biological coefficient of variation (BCV) for pairwise comparison of samples. Samples are represented graphically in two dimensions (BCV distance 1 on the x-axis and BCV distance 2 on the y-axis).

As genotype was determined to be a major factor affecting gene expression profiles, most analyses hereafter assess transcriptome data for the two genotypes separately to focus on the effects of ploidy without the confounding effects of genotype on the transcriptome.

3.3.2.3 Transcriptional profiles between ploidy levels were strongly correlated

Linear regression plots of normalized gene expression counts between $2x_{\text{-colchicine}}$ and $4x_{\text{-colchicine}}$, $2x_{\text{-colchicine}}$ and $8x_{\text{-colchicine}}$, and $4x_{\text{-colchicine}}$ and $8x_{\text{-colchicine}}$ for both genotypes showed strong positive correlation between the ploidy pairs for both genotypes (Table 3.5). From the regression plots (Figures 3.13-3.15), it can be seen that the majority of points cluster very tightly around the 1:1 line (i.e. slope=1 i.e. no relative difference in transcription between the ploidy pairs) and the trendlines (linear regression line of best fit) for the data points in all ploidy comparisons were close to the 1:1 line. However, despite the trendlines appearing to be virtually superimposed on the 1:1 line, the

deviation from a 1:1 ratio, albeit minor, was significant (summarised in Table 3.5 and represented in Figures 3.13-3.15). The slope values (m) for the regression lines suggested that the deviation in tetraploid transcription profile in relation to the diploid transcription profile was greater for Sf when compared to Ler, and that of octoploids compared to diploids was greater for Ler when compared to Sf (Table 3.5)

From the linear regression plots, it was also observed that Sf polyploids had more genes that differed in expression (>2 fold) per gene in comparison to Ler for diploid-tetraploid and diploid-octoploid gene expression comparisons (Figures 3.13-3.14). We also found that as the difference between the ploidy comparisons increased from diploid v/s tetraploid (difference of 2), to tetraploid v/s octoploid (difference of 4), to diploid v/s octoploid (difference of 6), the data points on the regression plots were farther away from the 1:1 trendline, and the overall correlation in gene expression (per gene) decreased with increasing difference in ploidy between the ploidy levels being compared (Table 3.5). Diploid v/s tetraploid gene expression comparisons had the strongest correlation, followed by tetraploid v/s octoploid, and diploids v/s octoploids had a relatively weaker correlation (Table 3.5).

Table 3.5 | Summary of R-squared (R^2) values, slopes (m) and intercepts (c) obtained for the regression lines when tested for deviation from the hypotheses (R^2=0; and m=1). N corresponds to the number of genes in the test.

R function call: sma(y ~ x, log="xy", slope.test=1, method="SMA", alpha=0.05, multcompmethod="adjusted")

X	Y	N	R^2	p-value (R^2)	m	p-value (m)	c
$Sf2x_{-colchicine}$	$Sf4x_{-colchicine}$	17422	0.883	<2.22E-16	0.98	1.35E-14	0.047
$Sf2x_{-colchicine}$	$Sf8x_{-colchicine}$	17421	0.828	<2.22E-16	1.019	2.39E-09	-0.02
$Sf4x_{-colchicine}$	$Sf8x_{-colchicine}$	17419	0.857	<2.22E-16	1.04	<2.22E-16	-0.07
$Ler2x_{-colchicine}$	$Ler4x_{-colchicine}$	17821	0.913	<2.22E-16	0.992	3.84E-04	0.002
$Ler2x_{-colchicine}$	$Ler8x_{-colchicine}$	17817	0.853	<2.22E-16	1.044	<2.22E-16	-0.067
$Ler4x_{-colchicine}$	$Ler8x_{-colchicine}$	17819	0.897	<2.22E-16	1.05	<2.22E-16	-0.065

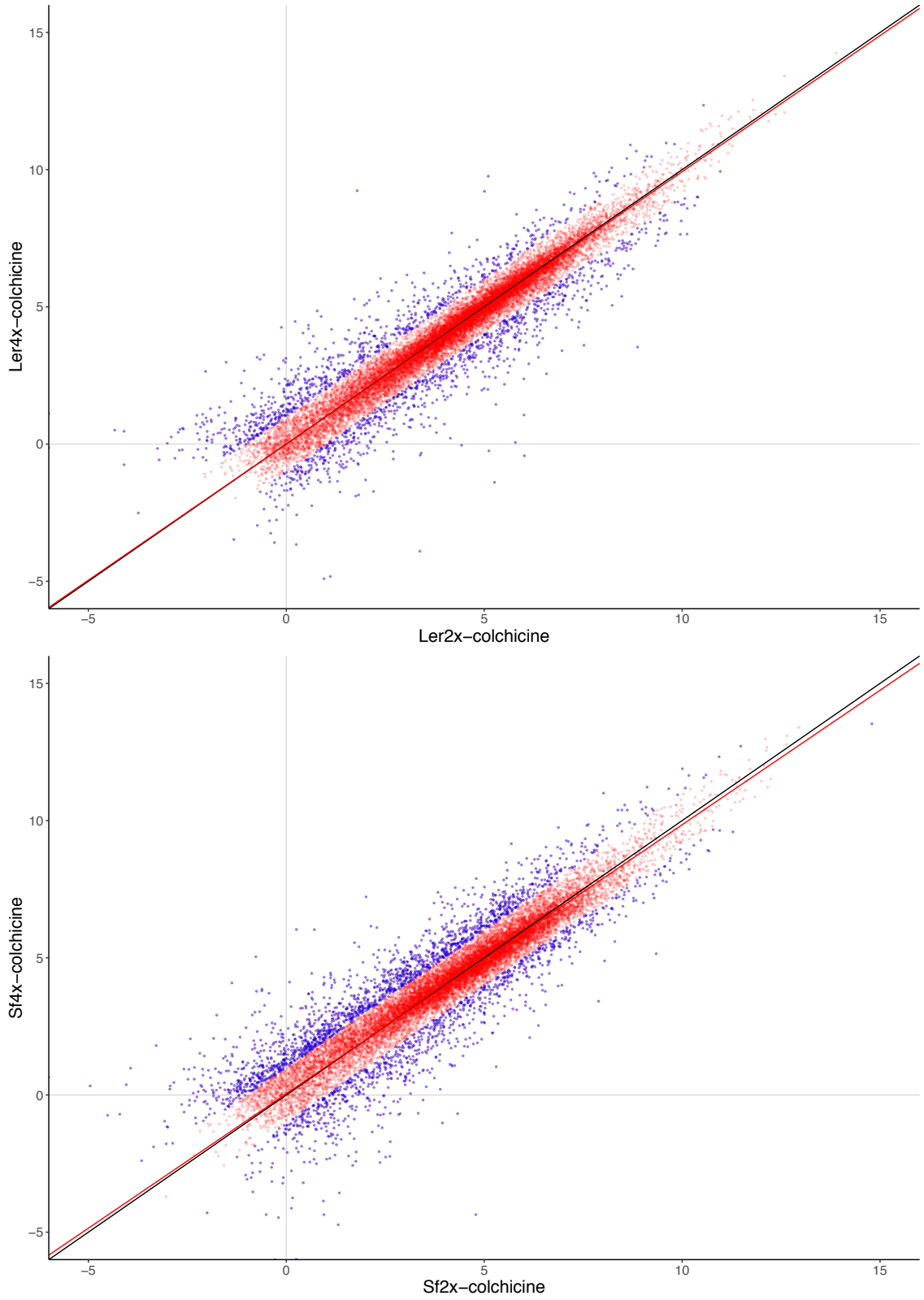


Figure 3.13 | Scatterplot of diploid v/s tetraploid gene expression for (A) Ler and (B) Sf. Black line = theoretical 1:1 line passing through origin; red line = linear regression line for the data; red dots = normalized gene expression counts (log transformed) for diploids (x-axis) and tetraploids (y-axis); blue dots = genes showing a >2 log fold-change in expression between diploids and tetraploids.

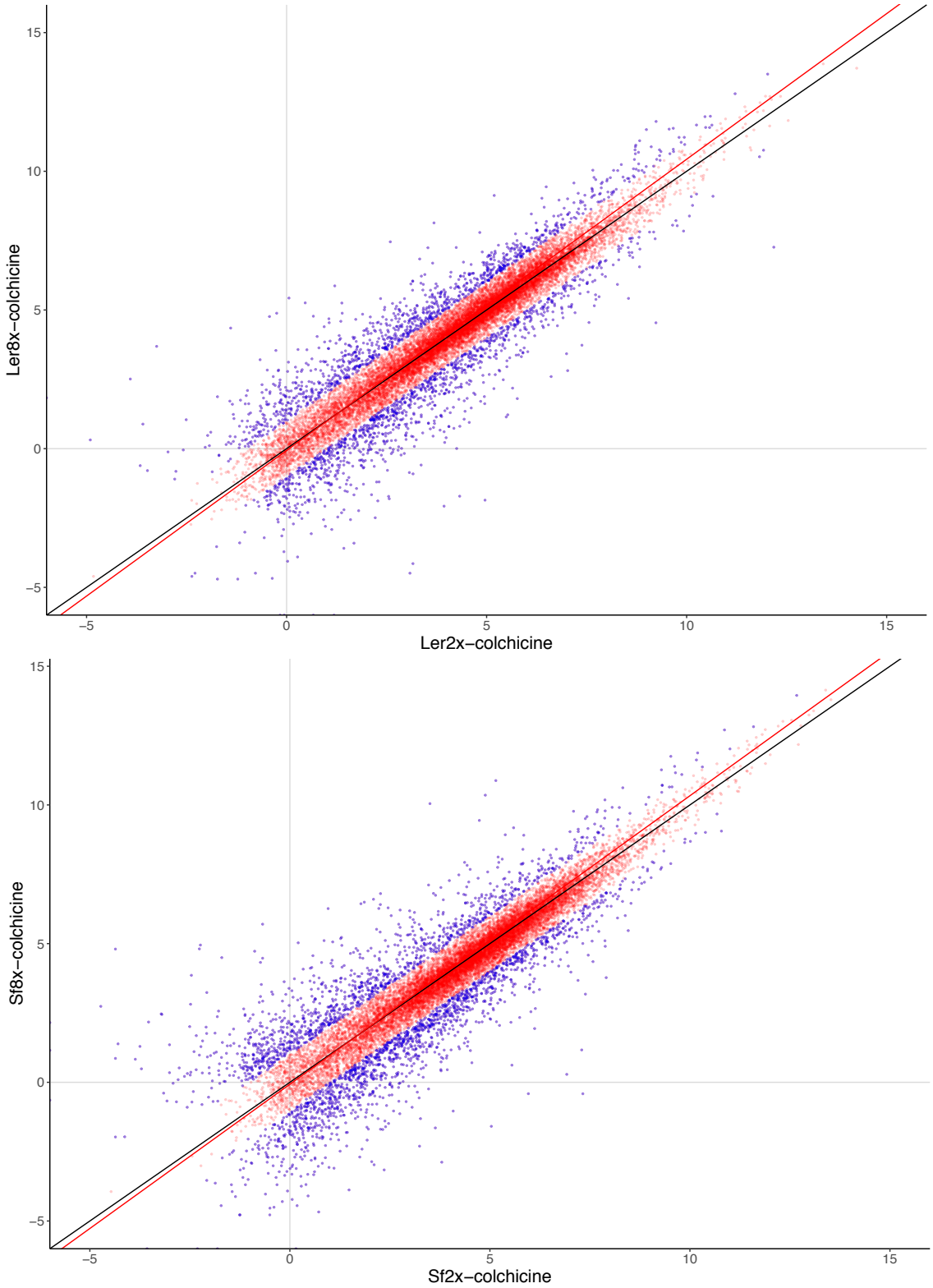


Figure 3.14 | Scatterplot of diploid v/s octoploid gene expression for (A) Ler and (B) Sf. Black line = theoretical 1:1 line passing through origin; red line = linear regression line for the data; red dots = normalized gene expression counts (log transformed) for diploids (x-axis) and octoploids (y-axis); blue dots = genes showing a >2 log fold-change in expression between diploids and tetraploids.

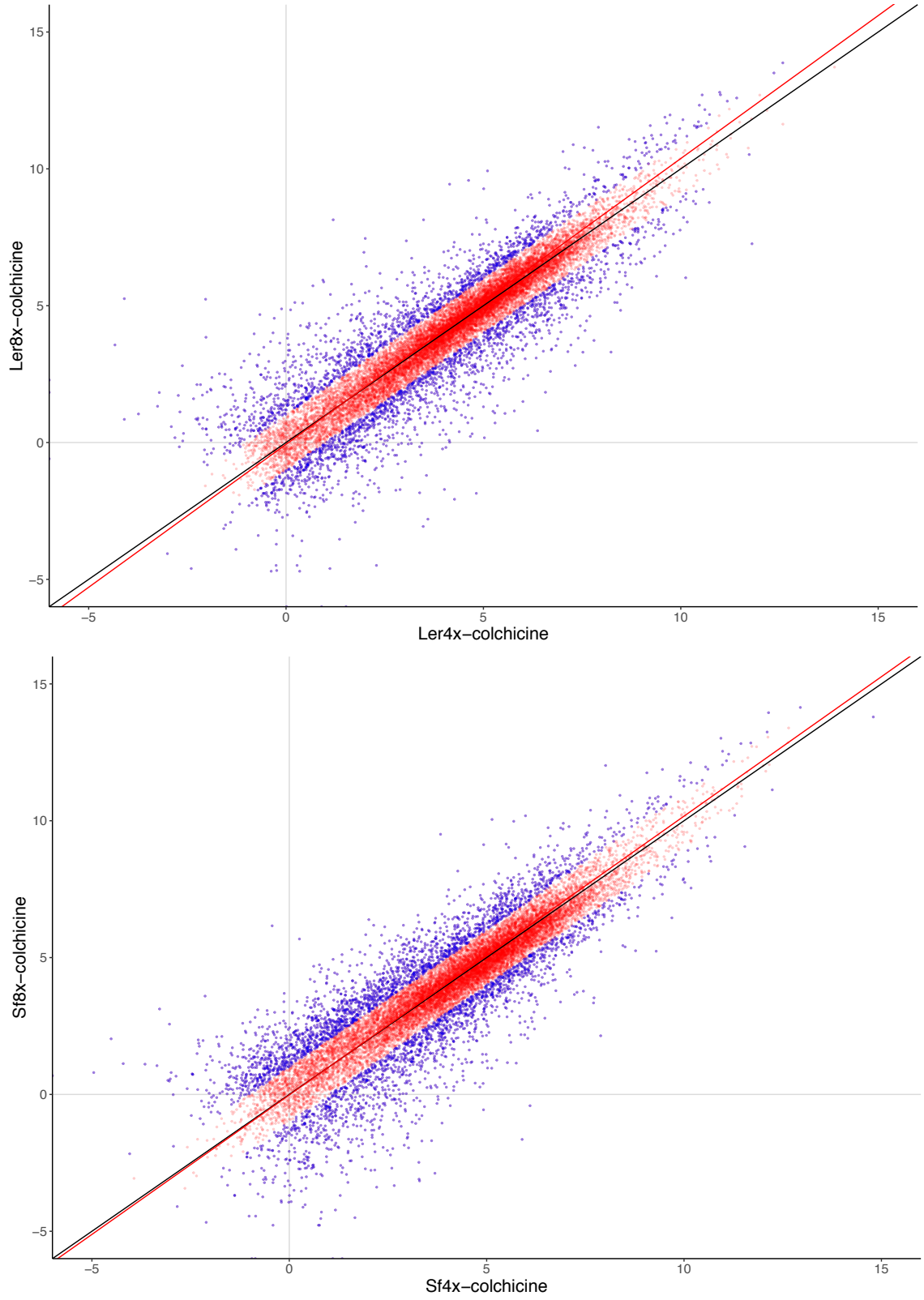


Figure 3.15 | Scatterplot of tetraploid v/s octoploid gene expression for (A) Ler and (B) Sf. Black line = theoretical 1:1 line passing through origin; red line = linear regression line for the data; red dots = normalized gene expression counts (log transformed) for tetraploids (x-axis) and octoploids (y-axis); blue dots = genes showing a >2 log fold-change in expression between diploids and tetraploids.

3.3.2.4 Gene expression differences between ploidies were uniform across the chromosomes.

In order to explore whether there was over-expression or under-expression in specific regions of the chromosome, we analysed gene expression changes as a function of the location on the chromosome. A representative plot for Ler gene expression data for Chromosome 1 can be seen in Figure 3.16 (plots for all chromosomes for both genotypes are available in the Supplementary Data 7.4-7.12). The distribution of points along the x-axis (genomic coordinates of the chromosome) and y-axis (log_2-Fold Change) was homogenous along the entire length of the chromosome, except around the centromere region (this is because there are fewer genes at the centromere, and not because of an expression bias). This suggested that there was no evident pattern of expression change corresponding to specific regions on the chromosome that could be attributed to deletions or rearrangements of large DNA fragments in the polyploids.

3.3.2.5 The majority of the genes showed no change in expression with increasing ploidy

The gene expression data for colchicine treated plants averaged across the three biological replicates for each genotype and ploidy level suggested that there was no change in gene expression across the three ploidies for 78% of the genes (defining genes that had expression changes as those that had a ≥2-fold change in expression between the ploidy levels). About 17% of the genes showed a ≥2-fold change in expression between diploids and tetraploids only, or between tetraploids and octoploids only. Within these, the proportion that change between tetraploids and octoploids was higher (10%) than those between diploids and tetraploids (7%). Interestingly, there were more genes that switched their direction of expression as they move from diploids to tetraploids to octoploids (3%) and only 1% of the genes had a linear increase or decrease in relative gene expression across ploidy levels.

When comparing between genotypes, we found that overall, Sf had a higher

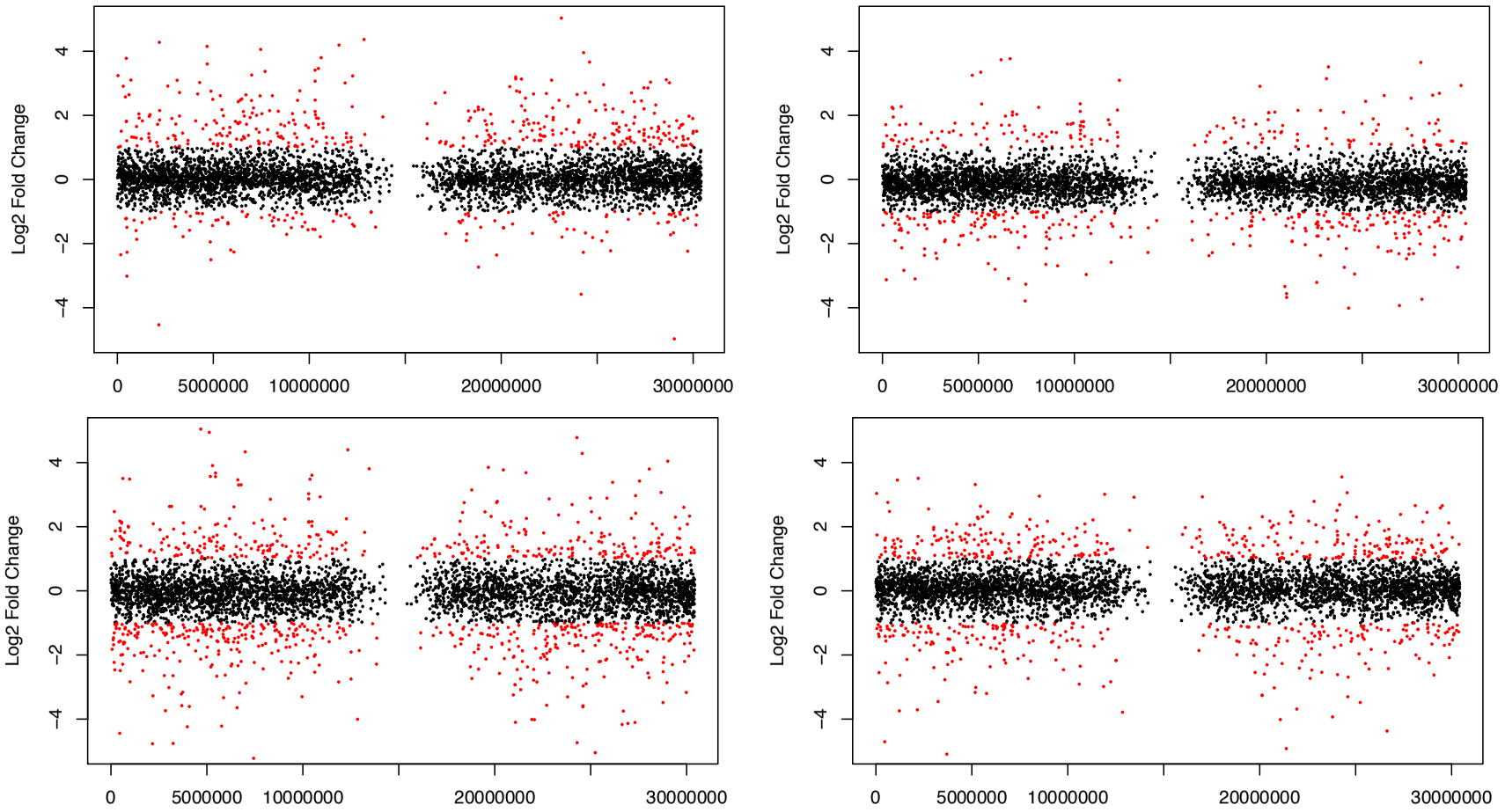


Figure 3.16 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) across Chromosome 1 (genomic coordinates on the x-axis). (A) Ler2x-colchicine to Ler2x-untreated (B) Ler4x-colchicine to Ler2x-colchicine (C) Ler8x-colchicine to Ler2x-colchicine (D) Ler8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.

number of genes that showed a ≥2-fold change in expression when compared to Ler. Sf also had more variation in genes that had no change between diploids and tetraploids and increase between tetraploids and octoploids (773 genes), and those that decrease expression between tetraploids and octoploids (1004). Similar variation was also seen in Sf genes that switched the direction of expression (440 v/s 377), or showed a linear increase or decrease in expression (168 v/s 85), or those that increased or decreased expression between diploids and tetraploids and remained constant between tetraploids and octoploids (1066 v/s 446). In contrast, the numbers for Ler genes were more consistent in terms of the direction (increase or decrease) within each case (Figure 3.6).

2x - 4x - 8x	Expression	2x - 4x - 8x	Expression (2x-4x vs 4x-8x)	Ler	Sf
A	B No change (78.05%)	C	D No change - No change	E 14474	F 13067
	2x - 4x : No change 4x - 8x : Change (9.84%)		No change - Increase	816	773
			No change - Decrease	874	1004
	2x - 4x : Change 4x - 8x : No change (7.25%)		Increase - No change	501	1066
			Decrease - No change	541	446
	2x - 4x - 8x : Switch / Change in direction (3.24%)		Increase - Decrease	180	440
			Decrease - Increase	210	377
	2x - 4x - 8x : Linear increase or decrease (1.36%)		Linear decrease	113	168
			Linear increase	114	85

Figure 3.17 | Directionality of gene expression change across ploidy levels. (A-B) Direction of gene expression change categorised into five groups and the percentage of genes corresponding to each group. (C-D) Sub-categorisation of the direction of gene expression changes and the number Ler and Sf genes belonging to each sub-category (E-F). Blue lines: <2-fold change in expression; green lines: ≥2-fold increase in expression between diploids and tetraploids, or between tetraploids and octoploids; red lines: ≥2-fold decrease in expression between diploids and tetraploids, or between tetraploids and octoploids.

3.3.3 Correlation analysis revealed that a greater number of Sf genes were significantly correlated with ploidy in comparison to Ler genes.

We conducted correlation analysis on normalized gene expression data to specifically evaluate genes that showed a strong correlation in expression across the ploidies. The correlation coefficient and p-value thresholds beyond which genes were declared significantly correlated were determined by a bootstrapping approach. Since the effects of genotype and colchicine could not be accounted for and isolated in the correlation analysis, the analysis was run separately for each genotype and using data for colchicine treated lines only.

We found that 1402 genes were significantly correlated for Ler and 3876 genes were significantly correlated for Sf. Scatterplots of the top five positively and negatively correlated genes for each genotype are shown in Figure 3.18.

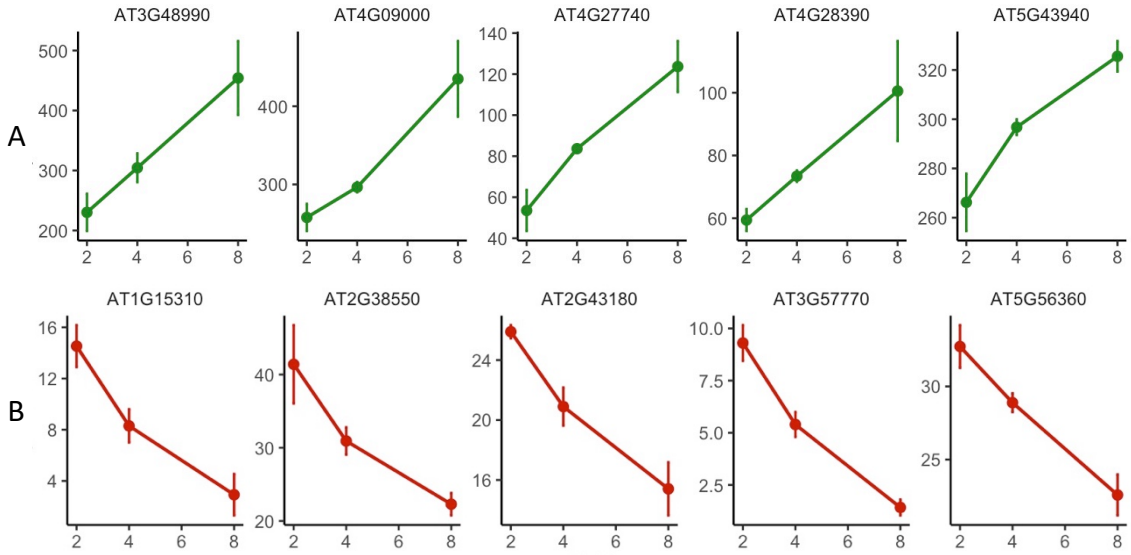


Figure 3.18 | Top positively and negatively correlated genes for Ler. Thresholds for significance and Spearman's rank correlation coefficient (ρ) were determined using a bootstrapping approach (bootstrap=10,000). P-values that fell below the 5th percentile of bootstrap p-values were considered significant. Rho-values (ρ) that were below the 5th percentile of bootstrapped values were considered negatively correlated and those above the 95th percentile were considered positively correlated. Panel A = top five significant positively correlated genes; panel B = top five significant negatively correlated genes; x-axis = ploidy level; y-axis = normalized gene expression counts; sub-plot titles = gene names.

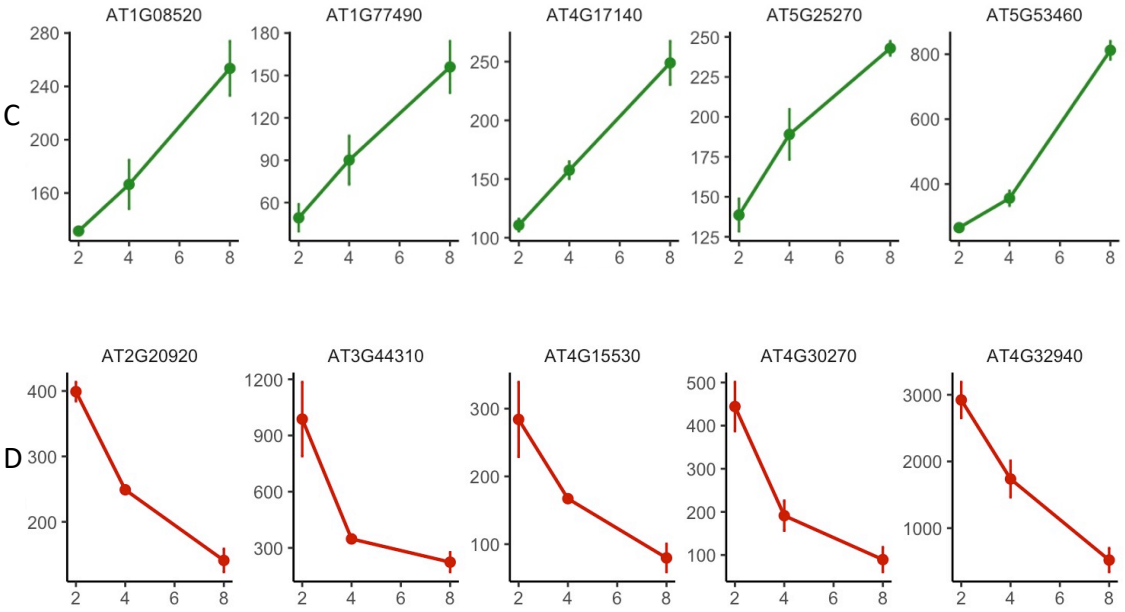


Figure 3.19 | Top positively and negatively correlated genes for Sf. Thresholds for significance and Spearman's rank correlation coefficient (ρ) were determined using a bootstrapping approach (bootstrap=10,000). P-values that fell below the 5th percentile of bootstrap p-values were considered significant. Rho-values (ρ) that were below the 5th percentile of bootstrapped values were considered negatively correlated and those above the 95th percentile were considered positively correlated. Panel A = top five significant positively correlated genes; panel B = top five significant negatively correlated genes; x-axis = ploidy level; y-axis = normalized gene expression counts; sub-plot titles = gene names.

3.3.4 Analysis of differentially expressed genes using EdgeR

3.3.4.1 Combined model (all samples): Genotype had the largest effect on the differential expression of genes between samples

Using the statistical package EdgeR, normalized gene expression data for 18592 genes were analysed. All 24 samples were considered within the same model that took into account the effects of genotype, colchicine treatment, ploidy, and genotype by ploidy interaction on gene expression. A gene was defined as differentially expressed if the False Discovery Rate (FDR) value (p-value corrected for multiple comparisons) was <0.05.

Differential expression analysis using a combined linear model including all factors showed that the genotype was predominantly responsible for differential expression of genes between the 24 samples (8002 genes; 43.04%). Furthermore, there were more genes that were differentially expressed as a result of genotype-by-ploidy interaction (1008 genes; 5.42%) than due to ploidy alone (473 genes; 2.54%). Colchicine appeared to have a relatively minor effect on gene expression (6 genes; 0.003%). In concurrence with our earlier findings (Figure 3.12), most of the genes determined to be differentially expressed were in response to differences between the genotypes.

In order to understand the functions of the identified DEGs, we conducted Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis on the DEG list for ploidy and genotype-by-ploidy interaction. There was significant enrichment for biological processes (BP) for both gene sets. For genes that were differentially expressed in response to ploidy, there was significant enrichment for response to light, response to sucrose, photorespiration, circadian rhythm, photosynthesis, response to karrikin, response to cold, and redox processes (or oxidation-reduction reactions i.e. chemical reactions involving electron transfers) (Figure 3.20). The KEGG pathways that were enriched included metabolic pathways for compounds like alanine, aspartate, glutamate, glyoxylate, dicarboxylate, fructose, mannose and carbon, carbon fixation, and for the biosynthesis of secondary metabolites (Figure 3.20). For genes that were differentially expressed in response to genotype-by-ploidy interaction, enriched GO terms included those for starch catabolic and biosynthetic processes, microtubule-based processes, circadian rhythm, phototransduction, response to karrikin, cold, abscisic and salt stress (Figure 3.21), and only one KEGG pathway – biosynthesis of secondary metabolites, was enriched (Figure 3.21).

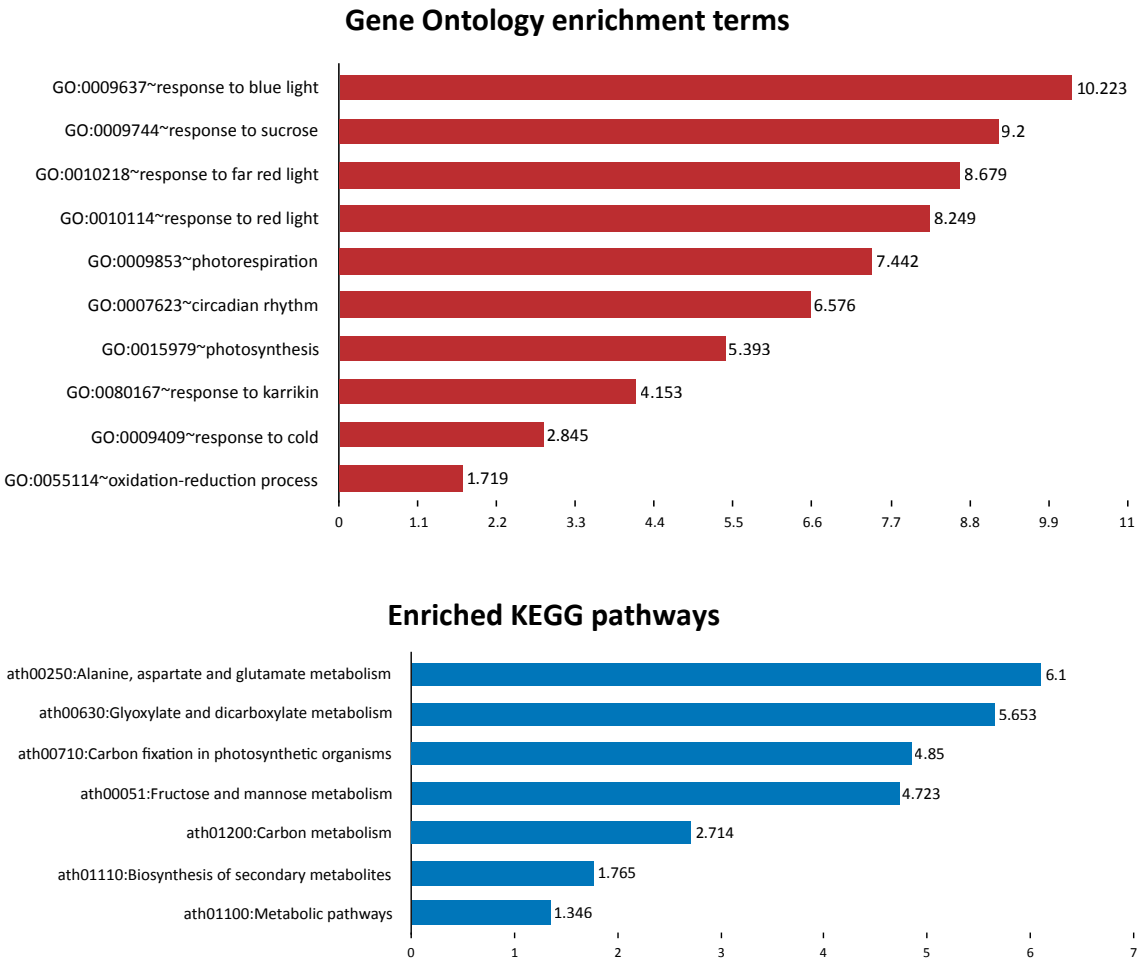


Figure 3.20 | Gene Ontology and KEGG pathway analysis of differentially expressed genes in response to ploidy. The functional annotation tool DAVID (v6.8; https://david.ncifcrf.gov) was used to identify Gene Ontology terms (red) and KEGG pathways (blue) that were significantly enriched for the differentially expressed genes in response to ploidy. P-values were adjusted using the Benjamini-Hochberg correction for multiple testing and values<0.05 were considered significant. X-axis and numbers to the right of the bars indicate fold change, y-axis indicates the GO term number and category title and KEGG pathway codes and names for the top and bottom plot respectively.

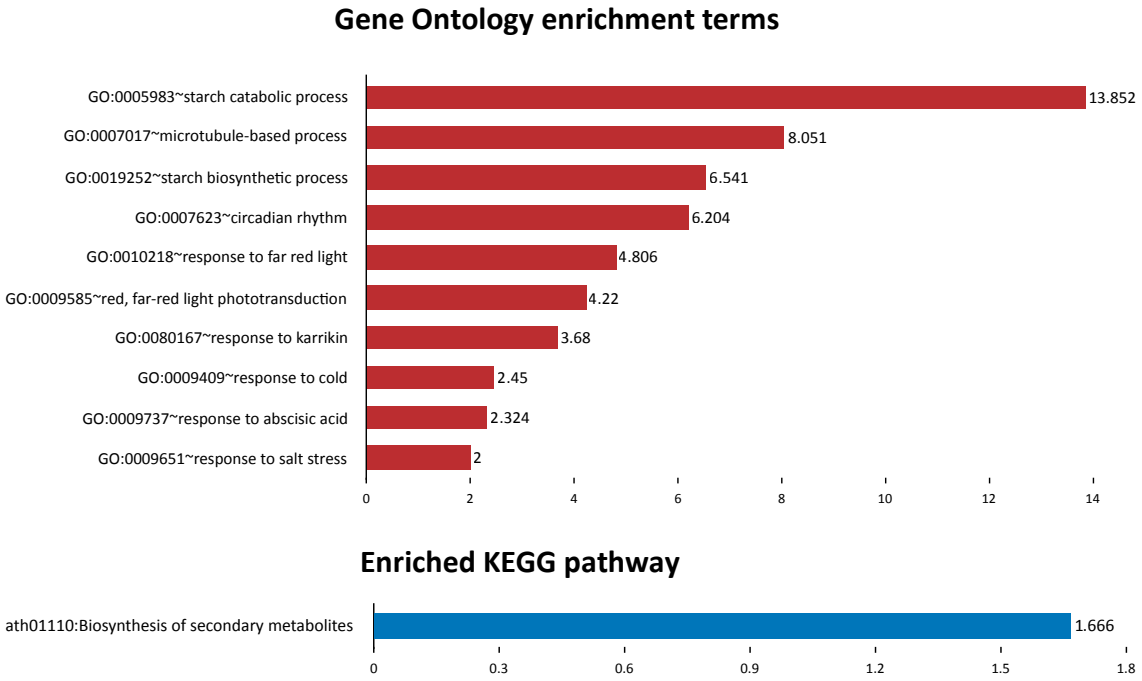


Figure 3.21 | Gene Ontology and KEGG pathway analysis of differentially expressed genes in response to genotype-by-ploidy interaction. The functional annotation tool DAVID (v6.8; https://david.ncifcrf.gov) was used to identify Gene Ontology terms (red) and KEGG pathways (blue) that were significantly enriched for the differentially expressed genes in response to genotype-by-ploidy interaction. P-values were adjusted using the Benjamini-Hochberg correction for multiple testing and values<0.05 were considered significant. X-axis and numbers to the right of the bars indicate fold change, y-axis indicates the GO term number and category title and KEGG pathway codes and names for the top and bottom plot respectively.

As most genes were differentially expressed in response to ploidy or genotype-by-ploidy interaction, additional analyses were conducted using data from each genotype independently to understand which genes are differentially expressed in response to ploidy within the specific genotypes and how these compare to those that were determined to be differentially expressed in response to genotype-by-ploidy interaction or ploidy in the combined analysis model.

3.3.4.2 Individual model: The identity of DEGs and their functional were non-overlapping for the two genotypes

A principal component analysis (PCA) conducted individually for each genotype showed distinct separation between diploids and octoploids, with tetraploids interspersed between the two along PC1 that accounted for 61.51% of the variation in gene expression for Ler and 63.71% for Sf (Figure 3.22-3.23). When comparing the factor loadings of the top 100 genes contributing towards the variation along PC1, 57 genes overlapped between Ler-0 and Sf-2 suggesting that these are likely to be genes that have a common response to ploidy for both genotypes. The remainder of the genes could either be those that are unrelated to ploidy or are genotype specific.

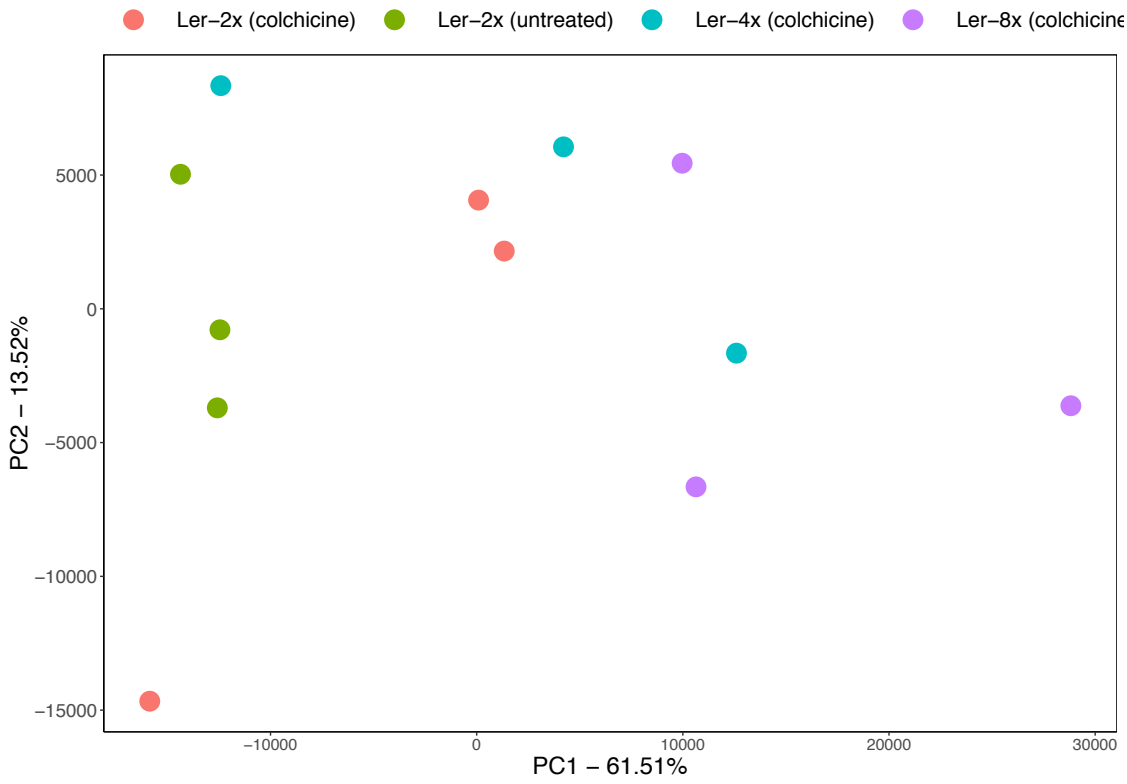


Figure 3.22 | Principal component analysis (PCA) of normalized gene expression counts for Ler samples. 17528 genes were analysed across 12 samples belonging to colchicine treated lines for three ploidy levels and a diploid untreated line (three biological replicates each). Principal component 1 (PC1; x-axis) accounted for 61.51% of the overall variance and principal component 2 (PC2; y-axis) accounted for 13.52% of the variance.

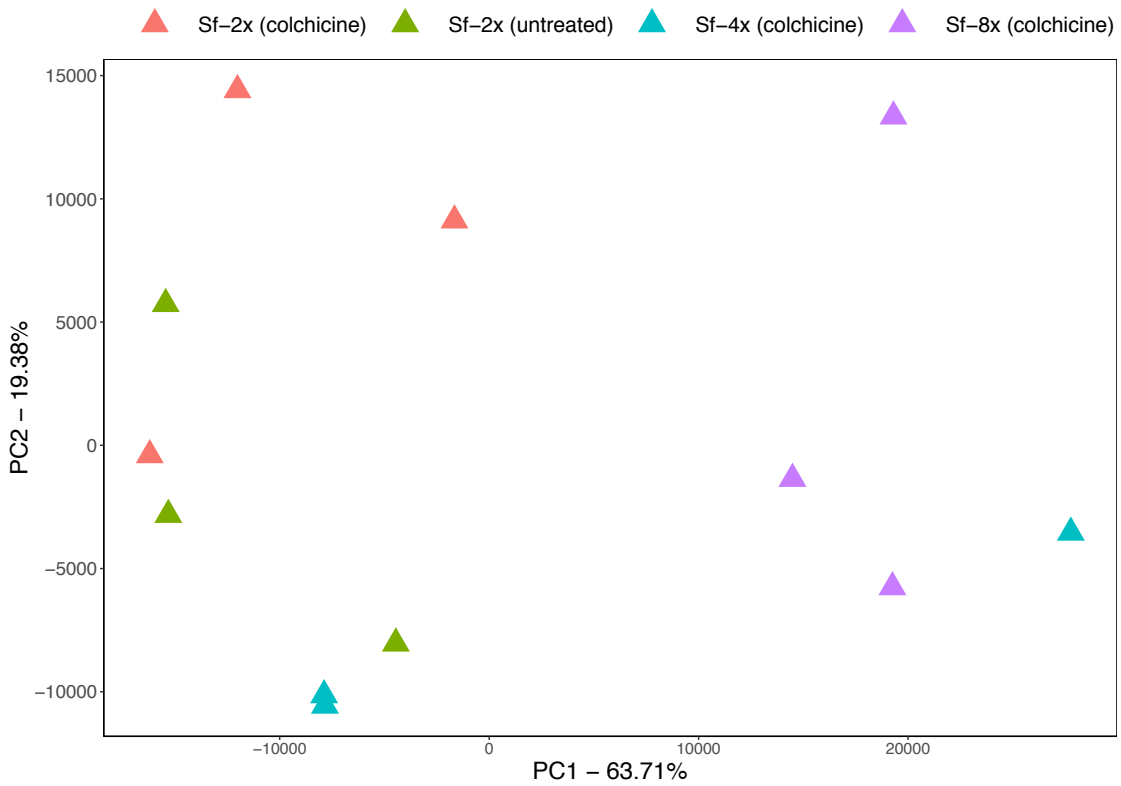


Figure 3.23 | Principal component analysis (PCA) of normalized gene expression counts for Sf samples. 17149 genes were analysed across 12 samples belonging to colchicine treated lines for three ploidy levels and a diploid untreated line (three biological replicates each). Principal component 1 (PC1; x-axis) accounted for 63.71% of the overall variance and principal component 2 (PC2; y-axis) accounted for 19.38% of the variance.

In order to determine genotype-specific gene expression responses to increased ploidy, we analysed the gene expression data using the statistical package EdgeR for the two genotypes independently using an additive model that took into account colchicine and ploidy effects. We observed a >10-fold difference in the number of DEG's obtained between the two genotypes in response to ploidy. Sf had 1988 DEGs in response to increased ploidy and Ler-0 had 149 DEGs. There was limited overlap between the genes obtained for Ler and Sf (29 genes), suggesting that majority of the DEGs obtained in response to ploidy from individual analyses were specific to that genotype. Furthermore, the genes that were differentially expressed for each genotype had different functional characterizations. DEGs in response to ploidy in Ler were enriched

for response to hydrogen peroxide, karrikin and cold, and for circadian rhythms (Figure 3.24). There was a >20-fold enrichment for circadian rhythm KEGG pathway genes (Figure 3.24). For Sf on the other hand, there was enrichment for metabolic and biosynthetic processes, circadian rhythm, cytokinin, and cell homeostasis (Figure 3.24). Enriched KEGG pathways included biosynthesis of steroids and secondary metabolites, and

Table 3.6 | Summary of the number of differentially expressed genes obtained for the three analysis models. DEGs were determined using the R package EdgeR. Model = analysis model; Indiv = individual analysis; samples = number of samples used for the analysis; colchicine, genotype and ploidy = number of DEGs obtained in response to each effect; GxP = number of DEGs obtained in response to genotype-by-ploidy interaction effects; 2-4 = number of DEGs between diploids and tetraploids; 4-8 = number of DEGs between tetraploids and octoploids; and 2-8 = number of DEGs between diploids and octoploids.

Model	Samples	Colchicine	Genotype	Ploidy	GxP	2-4	4-8	2-8
Combined	24	6	8002	473	1008	2	148	616
Indiv - Ler	12	0	NA	149	NA	0	133	91
Indiv - Sf	12	35	NA	1988	NA	176	718	1842

The observance of nearly double the number of genes in genotype-by-ploidy interaction (1008, in comparison to 473 ploidy-only genes) in the combined model, as well as the difference in the number and functional classification of genes that were differentially expressed in response to ploidy in the independent analyses for Ler and Sf reiterate that gene expression changes in response to increased ploidy are not only a function of ploidy increase but also of the genetic background of the plant.

A summary of the number of genes that overlap in the various analysis models are provided in Supplementary Data 7.13.

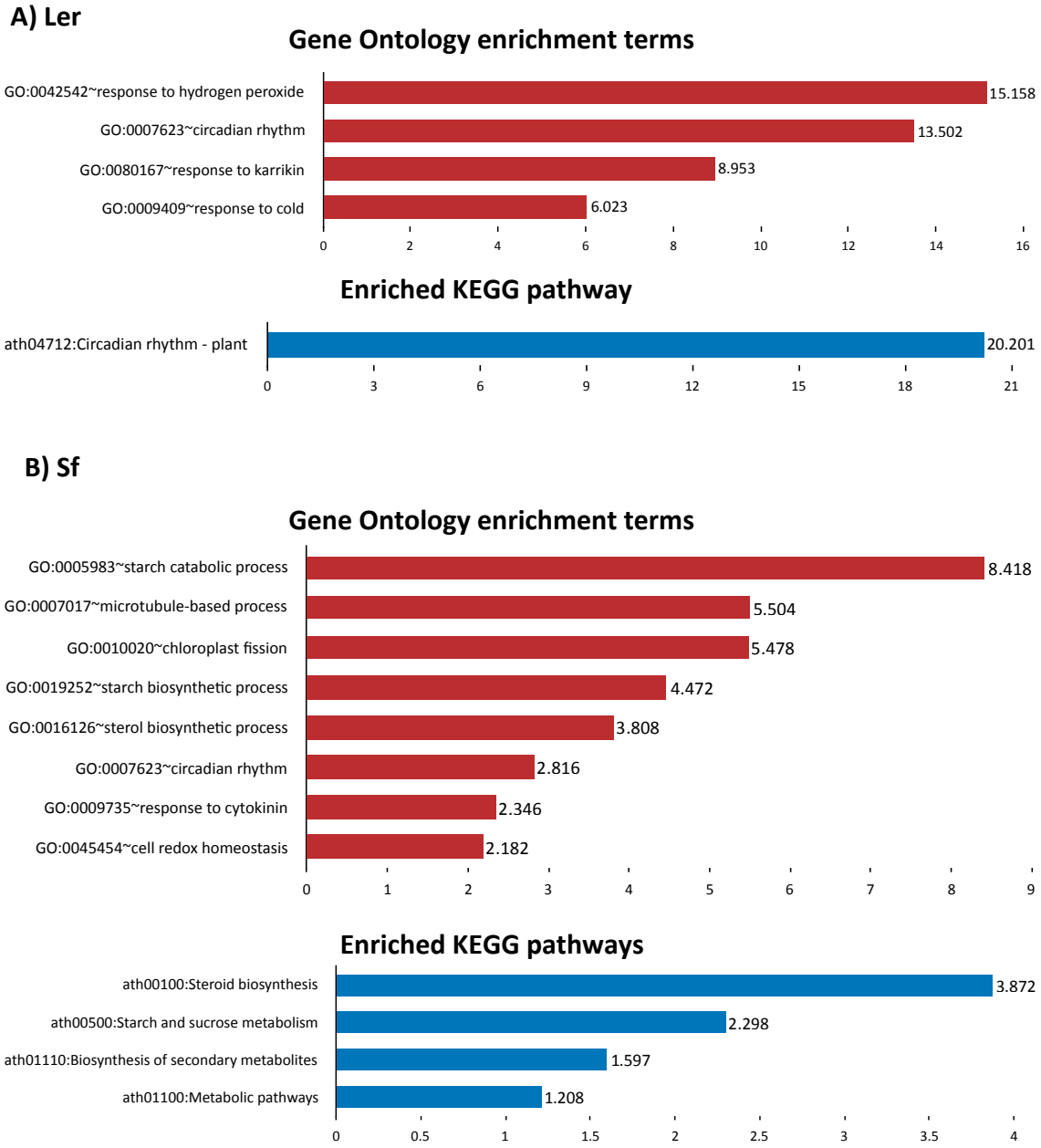


Figure 3.24 | Gene Ontology and KEGG pathway analysis of differentially expressed genes in response to ploidy for genotypes Ler and Sf. The functional annotation tool DAVID (v6.8; https://david.ncifcrf.gov) was used to identify Gene Ontology terms (red) and KEGG pathways (blue) that were significantly enriched for the differentially expressed genes in response to ploidy for (A) Ler and (B) Sf. P-values were adjusted using the Benjamini-Hochberg correction for multiple testing and values<0.05 were considered significant. X-axis and numbers to the right of the bars indicate fold change, y-axis indicates the GO term number and category title and KEGG pathway codes and names for the top and bottom plot respectively.

3.3.4.3 DEGs and SCGs included genes that are involved in determining various phenotypic traits

DEGs from all three analysis models i.e. combined, individual – Ler, and individual – Sf; and SCGs from Ler and Sf correlation analyses were scanned to check if their functions could explain any of the phenotypic traits previously observed (section 3.3.1). A summary of genes identified that could be involved in stomatal processes (guard cell differentiation, stomatal complex morphogenesis, stomatal complex development), trichome processes (trichome differentiation, trichome morphogenesis and trichome branching), flowering (photoperiodism, flowering time, regulation of photoperiodism), and leaf development (leaf morphogenesis, leaf shaping, leaf formation, leaf vascular tissue pattern formation, leaf development, regulation of leaf development) are listed in Tables 3.7-3.10.

Overall, within all the analyses of DEGs and SCGs, we found 18 genes were associated with stomatal processes (Table 3.7). Of the 18, only one gene was common between Sf and Ler. We found 18 genes that were associated with trichome processes, again, only one gene overlapped between Sf and Ler (Table 3.8). 38 genes were known to have functions related to the regulation of flowering time; five genes overlapped between Sf and Ler (Table 3.9). 101 genes had known functions related to leaf growth and development; 9 genes overlapped between Sf and Ler (Table 3.10). Although both genotypes elicit similar phenotypic responses to increased ploidy, the limited overlap between the genes obtained suggests that the phenotypic traits are likely to be modulated by different sets of genes.

Table 3.7 | DEGs and SCGs that could be involved in determination of stomatal phenotype. Gene = TAIR gene ID; analysis model 'Ler' includes DEGs and SCGs from independent Ler differential expression analysis and correlation analysis; analysis model 'Sf' includes DEGs and SCGs from independent Sf differential expression analysis and correlation analysis; analysis model 'Combined' includes DEGs obtained in response to ploidy (DEG_P) and genotype-by-ploidy interaction (DEG_I). Numeric values indicate the normalized gene expression counts for Ler and Sf samples respectively, averaged across the three biological replicates. Green and red numbers indicate upregulation and downregulation with respect to its previous lower ploidy level within the same genotype.

Gene	Analysis Model			Ler Normalized Gene Expression			Sf Normalized Gene Expression		
	Ler	Sf	Combined	2x-colchicine	4x-colchicine	8x-colchicine	2x-colchicine	4x-colchicine	8x-colchicine
AT1G13260	SCG			1736.66	2288.43	3189.19	519.41	609.10	1892.44
AT1G25250	SCG			57.15	36.66	24.89	92.89	54.10	47.75
AT1G32640	SCG			4630.18	5451.62	2041.25	5656.31	2980.15	3376.65
AT1G33240		SCG		1245.15	1202.49	2165.97	1613.44	2317.99	2129.31
AT1G49040		SCG		907.33	827.02	847.58	758.44	592.94	732.92
AT1G70410		SCG		1684.21	1163.09	1863.61	1732.10	2202.54	2513.51
AT2G26330		DEG		156.37	215.61	227.55	54.45	148.16	60.40
AT2G46410		SCG,DEG		33.20	30.34	30.37	48.76	69.49	31.04
AT3G01500	DEG	SCG,DEG	DEG_P	298.94	1553.26	4052.37	2212.89	1247.05	4632.22
AT3G22780		SCG		140.20	170.81	131.84	145.45	119.39	183.83
AT3G22820	SCG			18.53	28.13	40.36	41.43	62.60	41.68
AT3G24140		DEG		72.48	113.92	89.90	50.55	78.50	76.59
AT3G48750	SCG			615.82	655.18	985.76	839.47	1193.20	770.99
AT3G57230	SCG			439.40	544.48	662.33	605.82	616.96	610.18
AT4G12970		SCG		0.97	5.02	7.20	43.02	46.33	73.40
AT4G34160		SCG,DEG		21.19	33.71	37.66	19.60	17.77	30.45
AT5G05620		SCG,DEG		57.00	54.28	27.15	50.18	44.68	62.70
AT5G10440	SCG			16.19	18.40	33.52	11.92	7.50	15.49

Table 3.8 | DEGs and SCGs that could be involved in determination of trichome phenotype. Gene = TAIR gene ID; analysis model 'Ler' includes DEGs and SCGs from independent Ler differential expression analysis and correlation analysis; analysis model 'Sf' includes DEGs and SCGs from independent Sf differential expression analysis and correlation analysis; analysis model 'Combined' includes DEGs obtained in response to ploidy (DEG_P) and genotype-by-ploidy interaction (DEG_I). Numeric values indicate the normalized gene expression counts for Ler and Sf samples respectively, averaged across the three biological replicates. Green and red numbers indicate upregulation and downregulation with respect to its previous lower ploidy level within the same genotype.

Gene	Analysis Model			Ler Normalized Gene Expression			Sf Normalized Gene Expression		
	Ler	Sf	Combined	$2x_{\text{-colchicine}}$	$4x_{\text{-colchicine}}$	$8x_{\text{-colchicine}}$	$2x_{\text{-colchicine}}$	$4x_{\text{-colchicine}}$	$8x_{\text{-colchicine}}$
AT1G13180		SCG		245.35	213.48	189.13	306.63	332.86	415.83
AT1G33240		SCG		1245.15	1202.49	2165.97	1613.44	2317.99	2129.31
AT1G56580		DEG		159.14	146.49	155.02	159.42	367.98	124.46
AT1G66740		SCG		211.29	239.58	278.78	253.48	347.95	274.42
AT1G80350		SCG,DEG		192.39	188.59	209.76	182.55	161.53	186.96
AT2G27600	SCG			2241.73	2479.08	3201.76	1824.74	1958.32	1781.01
AT2G38440		SCG		754.52	840.82	1059.37	696.94	669.60	764.11
AT3G20780		SCG		66.03	61.51	42.84	82.69	64.99	93.37
AT4G00100		SCG		1304.69	1205.68	1251.90	1601.84	1849.38	1678.46
AT4G15415	SCG			592.96	641.70	777.74	603.02	705.85	639.09
AT4G22910		SCG,DEG		69.89	60.83	61.73	69.31	52.12	84.10
AT4G38600		SCG		3399.46	3856.91	3614.45	2442.29	2238.91	2415.89
AT5G18410		SCG,DEG		358.46	401.35	355.91	455.23	409.26	466.73
AT5G24310	SCG	SCG,DEG		260.38	247.95	162.86	234.21	287.81	213.85
AT5G24630	SCG			128.44	104.85	60.13	75.13	59.71	70.96
AT5G40330			DEG_P	2.27	4.52	23.84	13.20	53.18	19.37
AT5G53200			DEG_I	26.30	8.49	36.79	3.03	3.06	4.65
AT5G64930	SCG		DEG_I	345.99	405.42	501.87	293.64	360.64	240.54

Table 3.9 | DEGs and SCGs that could be involved in determination of flowering time. Gene = TAIR gene ID; analysis model 'Ler' includes DEGs and SCGs from independent Ler differential expression analysis and correlation analysis; analysis model 'Sf' includes DEGs and SCGs from independent Sf differential expression analysis and correlation analysis; analysis model 'Combined' includes DEGs obtained in response to ploidy (DEG_P) and genotype-by-ploidy interaction (DEG_I). Numeric values indicate the normalized gene expression counts for Ler and Sf samples respectively, averaged across the three biological replicates. Green and red numbers indicate upregulation and downregulation with respect to its previous lower ploidy level within the same genotype.

Gene	Analysis Model			Ler Normalized Gene Expression			Sf Normalized Gene Expression		
	Ler	Sf	Combined	2x-colchicine	4x-colchicine	8x-colchicine	2x-colchicine	4x-colchicine	8x-colchicine
AT1G01060	DEG		DEG_I, DEG_P	582.99	216.61	1457.21	41.66	29.67	60.66
AT1G22770	DEG		DEG_I, DEG_P	1397.16	2074.53	517.21	1703.00	1498.35	1659.20
AT1G51140		SCG		827.24	431.47	767.40	2737.58	1098.08	797.56
AT1G65480		SCG		49.38	103.41	27.07	39.40	22.79	43.30
AT2G03500		SCG,DEG	DEG_I	74.23	79.07	25.40	30.21	18.78	83.81
AT2G06255		SCG,DEG	DEG_I	151.38	206.16	187.92	190.08	291.58	186.91
AT2G21070		SCG		60.46	56.04	58.86	84.31	88.92	111.84
AT2G22370	SCG			155.77	139.15	127.63	138.42	137.69	150.00
AT2G29950		SCG,DEG		37.89	46.55	13.88	36.10	38.08	30.86
AT2G40080	DEG	SCG	DEG_I, DEG_P	181.68	26.49	1.05	406.10	294.24	584.51
AT2G42400		SCG		492.75	512.85	751.77	559.10	682.41	585.73
AT2G44680		SCG		726.56	820.45	826.48	816.83	1015.20	719.73
AT2G46340		SCG,DEG	DEG_I	1191.00	1038.75	1167.56	551.58	711.99	588.42
AT3G07650			DEG_I, DEG_P	139.99	119.23	11.80	2261.92	2605.68	2602.04
AT3G09600		SCG,DEG	DEG_I	372.74	248.69	648.10	42.81	84.66	45.15
AT3G10480	SCG	SCG,DEG		816.59	884.98	648.38	743.96	975.11	745.25
AT3G10490		SCG		933.81	1095.40	896.72	1102.97	1273.29	899.15
AT3G22231	DEG			1466.96	630.87	9692.53	593.94	391.42	2785.28
AT3G26640	SCG	SCG		194.98	246.64	300.12	201.77	210.34	254.78
AT3G57230	SCG			439.40	544.48	662.33	605.82	616.96	610.18
AT3G57390		SCG		202.38	206.87	265.21	211.68	218.50	178.45
AT4G03110	SCG			848.04	1108.55	2336.94	950.94	990.12	1497.49
AT4G26150		SCG	DEG_P	98.10	106.60	423.26	215.12	132.94	281.84
AT4G30200	SCG	SCG	DEG_I	892.26	1454.62	1934.96	2822.82	3511.76	1551.00
AT4G33980			DEG_I	345.53	273.95	38.25	1786.78	1482.10	2786.89
AT4G35900	SCG			123.33	141.52	68.23	77.51	71.96	165.74
AT4G37900		SCG,DEG	DEG_I	501.56	686.06	582.51	26.80	32.92	26.01
AT4G39100	SCG	SCG		884.06	1099.36	1287.02	1418.05	1465.01	1190.26
AT5G04240		SCG		492.59	590.34	591.06	605.69	508.98	622.20
AT5G06850		SCG,DEG	DEG_I	103.57	198.68	75.89	47.78	29.93	112.01
AT5G09530		DEG		222.17	48.90	4.70	381.36	60.66	297.59
AT5G17690	SCG			469.14	490.11	720.10	564.89	539.08	562.27
AT5G44160		SCG		42.62	59.10	35.10	17.94	14.91	29.91
AT5G46210		SCG		1757.77	2064.11	1946.12	1849.19	1841.86	1872.21
AT5G48300		SCG,DEG	DEG_I	2691.14	1322.09	2878.49	1503.54	967.20	2870.52
AT5G63980		SCG		1366.92	1170.31	861.86	1174.23	1201.95	1052.34
AT5G64930	SCG		DEG_I	345.99	405.42	501.87	293.64	360.64	240.54
AT5G65060			DEG_P	42.81	155.87	51.88	120.55	85.99	115.27

Table 3.10 | DEGs and SCGs that could be involved in determination of leaf development or morphology. Gene = TAIR gene ID; analysis model 'Ler' includes DEGs and SCGs from independent Ler differential expression analysis and correlation analysis; analysis model 'Sf' includes DEGs and SCGs from independent Sf differential expression analysis and correlation analysis; analysis model 'Combined' includes DEGs obtained in response to ploidy (DEG_P) and genotype-by-ploidy interaction (DEG_I). Numeric values indicate the normalized gene expression counts for Ler and Sf samples respectively, averaged across the three biological replicates. Green and red numbers indicate upregulation and downregulation with respect to its previous lower ploidy level within the same genotype.

Gene	Analysis Model			Ler Normalized Gene Expression			Sf Normalized Gene Expression		
	Ler	Sf	Combined	$2x_{\text{-colchicine}}$	$4x_{\text{-colchicine}}$	$8x_{\text{-colchicine}}$	$2x_{\text{-colchicine}}$	$4x_{\text{-colchicine}}$	$8x_{\text{-colchicine}}$
AT4G02030		SCG		528.46	536.67	637.79	469.72	488.52	481.09
AT4G02260		DEG		1315.40	1162.68	1267.99	1562.10	1006.73	2173.94
AT4G04890		DEG		652.30	737.50	930.65	409.72	1268.50	474.86
AT4G10090		SCG		117.51	135.90	157.29	158.10	154.46	137.56
AT4G20910		SCG,DEG		56.16	57.24	52.91	37.20	14.36	48.93
AT4G23750		SCG,DEG		102.28	128.87	185.24	54.73	21.97	40.77
AT4G30520		SCG		102.74	124.24	90.39	46.45	51.75	76.92
AT4G30790	SCG			1039.10	1227.41	1647.25	964.86	923.36	825.80
AT4G32940	SCG	SCG,DEG		58145.88	62423.20	29724.53	73088.42	58682.43	34856.58
AT4G34540		SCG		336.21	358.30	446.21	384.38	426.47	395.49
AT4G36380		SCG,DEG		73.99	40.78	78.75	110.44	139.29	55.23
AT4G39400		SCG,DEG	DEG_P	1028.40	2212.42	2995.39	2181.42	1978.21	2837.78
AT5G01600		SCG,DEG		5469.57	5760.48	4394.58	16269.49	12334.72	2855.56
AT5G02760	SCG	DEG		4.78	19.20	25.31	5.70	100.55	3.54
AT5G04810		SCG		1187.05	1433.91	1318.57	3107.52	2272.99	3387.79
AT5G05620		SCG,DEG		57.00	54.28	27.15	50.18	44.68	62.70
AT5G08520		SCG		887.49	1630.32	1451.71	1390.06	1696.67	1687.22
AT5G10250		SCG		3.17	6.59	3.25	7.81	16.37	41.67
AT5G13300		SCG		695.48	901.50	617.11	865.10	758.11	803.88
AT5G16470		DEG		1893.66	2416.82	2314.81	3034.85	4681.51	2566.32
AT5G17290		SCG,DEG		631.87	828.42	763.90	1692.97	1913.94	1200.53
AT5G27150		SCG		1421.86	1957.87	1598.25	3172.03	3426.74	2026.93
AT5G35750		SCG,DEG		139.67	128.05	126.67	312.25	336.34	350.05
AT5G37600		SCG		22675.82	23551.03	13917.65	1717.56	1518.55	7582.14
AT5G39610		SCG,DEG		876.45	1370.29	941.05	1172.16	1535.10	502.04
AT5G39740		SCG		1948.64	1660.42	1656.94	2359.86	2518.32	2354.99
AT5G44160		SCG		42.62	59.10	35.10	17.94	14.91	29.91
AT5G45890			DEG_P	54.83	10395.25	804.33	6.36	1.07	0.97
AT5G46210		SCG		1757.77	2064.11	1946.12	1849.19	1841.86	1872.21
AT5G46700		SCG		25.14	32.71	17.01	17.62	24.64	20.16
AT5G51720		DEG		115.77	73.52	190.43	32.63	12.87	273.76
AT5G53400	SCG	SCG,DEG	DEG_P	561.84	398.29	307.63	656.78	677.82	412.93
AT5G53860		SCG		438.09	567.33	505.95	960.52	730.21	1080.72
AT5G55540		SCG,DEG		178.20	220.39	230.50	140.99	116.54	191.73
AT5G56030	SCG		DEG_P	6215.28	2409.79	1642.24	2910.89	2224.79	2377.54
AT5G60970		SCG		89.39	119.85	157.40	32.24	48.70	60.32
AT5G62165			DEG_P	5.75	4.01	256.73	19.48	16.00	8.59
AT5G64930	SCG		DEG_I	345.99	405.42	501.87	293.64	360.64	240.54
AT5G64940			DEG_P	1785.07	1575.16	2891.99	726.15	682.16	953.78
AT5G66870	SCG	DEG		26.00	37.62	9.62	16.72	16.05	7.10
AT5G67100		SCG,DEG		40.71	18.31	22.33	23.19	20.96	21.98
AT5G67420	SCG			1293.35	2672.20	3503.67	1934.31	4210.75	4463.25

Table continues on the next page…

Gene	Analysis Model			Ler Normalized Gene Expression			Sf Normalized Gene Expression		
	Ler	Sf	Combined	2x-colchicine	4x-colchicine	8x-colchicine	2x-colchicine	4x-colchicine	8x-colchicine
AT1G01160	SCG	SCG		1256.36	948.18	1026.00	1502.71	1523.25	1313.14
AT1G04010	SCG			199.22	151.22	107.50	232.07	184.97	236.57
AT1G05180		SCG		703.43	613.59	632.93	582.13	514.51	647.83
AT1G11130		SCG		64.43	25.34	20.30	16.20	13.48	26.64
AT1G11760		SCG		123.97	113.94	127.54	166.39	177.50	164.13
AT1G13260	SCG			1736.66	2288.43	3189.19	519.41	609.10	1892.44
AT1G14000	SCG			1133.95	1222.73	1440.66	1191.69	1248.94	1200.16
AT1G17020		SCG,DEG		179.64	284.90	230.30	117.81	88.21	52.24
AT1G19270		SCG		1129.88	1181.31	807.62	551.36	353.52	878.92
AT1G19850		SCG,DEG		80.62	104.81	124.30	60.02	43.32	67.35
AT1G20780		SCG		703.67	487.57	807.10	548.98	369.55	706.54
AT1G25250	SCG			57.15	36.66	24.89	92.89	54.10	47.75
AT1G27370		DEG		236.94	183.48	204.84	194.70	247.87	267.29
AT1G30210	SCG			456.43	458.94	347.56	494.95	634.84	472.50
AT1G32450		SCG,DEG		184.50	426.47	168.33	795.65	754.78	250.66
AT1G45050		SCG		351.21	308.76	353.45	460.15	606.16	397.37
AT1G48410		SCG		2493.00	2482.01	2423.03	1970.15	1858.58	1949.70
AT1G54130		SCG		3415.22	3746.39	3256.69	2495.07	2403.83	2050.13
AT1G64360	SCG, DEG			6295.31	6948.65	2053.74	3543.73	2282.56	10296.29
AT1G69490		SCG,DEG		3840.21	6757.51	4141.59	5162.32	4155.05	1259.30
AT1G73500		SCG		2488.05	1891.55	1933.45	2062.80	1671.51	1081.15
AT1G79440	DEG	SCG,DEG	DEG_P	1171.71	1610.10	415.02	1447.68	1171.46	1930.54
AT2G19450			DEG_I	338.06	422.96	62.06	1015.61	789.33	1174.39
AT2G20570		SCG		6890.23	6433.85	5988.75	6088.94	7732.79	5221.60
AT2G21050		SCG,DEG		68.74	63.55	129.76	21.70	27.00	34.02
AT2G21185			DEG_I	509.87	704.67	1171.94	510.55	939.93	439.49
AT2G21660			DEG_I	60145.51	80026.34	3538.43	135158.91	126347.42	172154.82
AT2G23380		SCG,DEG		36.11	32.99	32.39	32.83	32.70	41.45
AT2G23430		SCG,DEG		899.20	793.02	847.51	1252.50	1184.87	1036.95
AT2G26330		DEG		156.37	215.61	227.55	54.45	148.16	60.40
AT2G29760			DEG_I	78.23	43.26	40.24	79.54	55.55	98.37
AT2G32700		SCG		2423.09	2354.19	2236.69	2084.97	2022.18	2215.04
AT2G36120		SCG,DEG		44.50	68.86	135.69	261.20	153.83	574.30
AT2G40300		SCG,DEG		672.80	659.71	387.22	1290.30	1234.48	666.94
AT2G42580		SCG,DEG		342.54	518.21	992.37	602.19	918.36	799.76
AT2G43000		SCG,DEG		248.71	181.01	287.53	18.51	46.36	17.69
AT2G43570	SCG	SCG		966.10	1628.67	3771.89	550.26	518.83	314.35
AT2G45450		DEG		116.18	209.58	138.21	215.85	436.00	117.35
AT2G46240			DEG_I	305.32	342.97	146.50	327.74	217.23	347.05
AT2G47180			DEG_I	235.57	531.03	348.55	263.82	284.57	131.59
AT2G48120	SCG	DEG		332.64	235.33	183.89	518.48	331.69	688.10
AT3G01470		SCG		3278.53	2997.06	3848.76	3182.33	3772.02	2711.24
AT3G02150	SCG			559.40	1079.46	1125.33	224.72	233.92	279.85
AT3G07050		SCG		397.38	267.54	257.94	295.47	374.18	299.26
AT3G14172			DEG_I	99.79	101.97	49.11	255.49	230.21	303.94
AT3G14940		SCG,DEG		23.06	48.10	12.81	18.57	15.13	49.34
AT3G27010		SCG		361.69	213.26	214.06	278.32	284.84	225.79
AT3G44680		SCG		249.74	238.16	279.55	152.77	129.78	142.40
AT3G47450		DEG		606.17	376.45	418.59	719.79	433.37	893.16
AT3G49500		SCG		452.62	421.21	475.84	564.93	379.78	545.78
AT3G50660		DEG		53.40	100.79	148.91	190.72	344.51	171.85
AT3G52860	SCG			223.18	213.78	160.95	222.32	190.73	204.35
AT3G54720		SCG,DEG		141.66	246.73	144.66	360.59	228.69	459.95
AT3G56090	SCG	SCG,DEG	DEG_P	408.66	267.59	39.80	623.19	394.96	501.66
AT3G62080		SCG		223.13	217.64	203.81	277.35	254.84	273.58
AT3G63300		SCG		48.01	37.63	38.43	28.97	18.25	23.45
AT3G63530		SCG		45.84	45.86	36.97	38.49	18.59	42.78
AT4G00100		SCG		1304.69	1205.68	1251.90	1601.84	1849.38	1678.46
AT4G01540		DEG		429.41	273.78	189.68	356.53	110.48	405.08

3.4 DISCUSSION

3.4.1 The relative expression levels of most genes are constant across ploidy levels

Although whole genome duplication is a common phenomenon, much remains unknown about how duplicated genes and genomes function in the initial stages of polyploidization. In the past, gene expression studies in neoautopolyploids have reported limited transcriptional differences between diploids and tetraploids (Albertin *et al.*, 2005; Wang *et al.*, 2006; Stupar *et al.*, 2007; Riddle *et al.*, 2010; Allario *et al.*, 2011; Gao *et al.*, 2016; Yin *et al.*, 2018), suggesting that gene expression in autopolyploids typically scales with ploidy for most genes. These observations are not surprising, given that large changes in sequence composition immediately after whole genome duplication are unexpected, especially in autopolyploids. However, the increase in phenotypic variation, biotic, and abiotic stress tolerance often observed in polyploids may suggest otherwise.

We observed that the linear regression lines for each ploidy-pair comparison showed only a minor deviation from the theoretical 1:1 line, with most points tightly clustered around it, suggesting that the relative expression levels between ploidy pairs were in a 1:1 proportion for most genes. Overall, 78% of the genes showed a <2-fold change in expression and only 1.36% of the genes had a >2-fold linear increase or decrease in expression across the ploidy levels. About 17% of the genes showed a >2-fold difference in expression between either diploid-tetraploid comparisons or tetraploid-octoploid comparisons. Interestingly, 3.24% of the genes switched their direction of expression with increasing ploidy (Figure 3.17). When logged fold-change in expression values were plotted along the chromosome co-ordinates, the distribution of genes that had a >2-fold difference in expression appeared to be uniformly distributed throughout the chromosome, suggesting that genes having large expression changes are not physically co-located on the chromosomes, thus ruling out the possibility of large chromosomal rearrangements or constraints to chromosome folding being reason behind the changes in expression observed in some genes.

3.4.2 Gene expression changes in response to increased ploidy in A. thaliana autopolyploids are genotype-specific

For both genotypes under study, the expression levels of most genes correlated for ploidy within the genotypes. Diploid-tetraploid comparisons showed the strongest

correlation with ploidy level, followed by tetraploid-octoploid comparisons, and diploid-octoploid comparisons. For each ploidy-pair comparison, Ler showed stronger correlation than Sf (Table 3.5) suggesting that there is more variation in expression between Sf genes at different ploidy levels than in Ler genes.

Comparative gene expression studies between diploids and neotetraploids in several species have usually been limited to a single genotype. Yu *et al.,* (2010) were the first to use multiple genotypes to study gene expression differences in response to ploidy. They used microarrays to study two genotypes of *A. thaliana* and reported that 476 genes were differentially expressed between diploid and neoautotetraploid Col, while only nine were differentially expressed between diploid and neoautotetraploid Ler. Our data also suggested similar genotype-specific patterns; however, we detected a larger number of differentially expressed genes in comparison, 1988 in Sf and 149 in Ler, likely due to the data from an additional ploidy level (octoploids) in the analysis model, relatively sensitive RNA-Seq methods, and the improved availability of analytical and statistical tools to detect differential expression over the past decade.

Exploiting natural genetic variation can be a valuable tool to understand how an increase in the genomic content (i.e. nucleotype) interacts with the genetic background (i.e. genotype) and why this results in varied in gene expression responses and consequently varied phenotypes. As the gene copy number in all neoautotetraploids is twice that of their respective diploids, genotype-specific changes in gene expression suggests that the expression of specific alleles or allele combinations is differentially regulated in some manner. The enrichment of specific GO terms and KEGG pathways further suggests that such allelic regulation is not stochastic.

In the combined analysis model, we found that 1008 genes were differentially expressed in response to genotype-by-ploidy interaction whereas only 473 were differentially expressed in response to ploidy only. This suggests that a number of genes between Ler and Sf have common responses to ploidy (genes involved in photosynthetic processes, metabolism of certain amino acids and carbohydrates, and in the biosynthesis of secondary metabolites) (Figure 3.20), but a larger number of genes have genotype-specific responses.

When the two genotypes were analysed individually to investigate gene expression differences specific to each genotype, we found that there was limited overlap (29 genes) between the genes that were differentially expressed between Ler

and Sf in response to polyploidy, suggesting that both genotypes have a distinct set of genes that are differentially expressed. Furthermore, we found that the differentially expressed genes in response to increased ploidy for the two genotypes were enriched for different functional categories. Ler showed an enrichment for the circadian rhythm pathway (Figure 3.24-A) whereas Sf predominantly shows an enrichment for steroid biosynthesis and metabolic pathways and to a lesser extent for circadian rhythm genes (Figure 3.24-B). Circadian rhythm genes are considered to be master regulators of metabolic and physiological pathways (Harmer *et al.*, 2000; Bendix *et al.*, 2015). In *Arabidopsis* allotetraploids, it has been shown that the circadian clock genes *CCA1* and *LHY* are epigenetically regulated by DNA methylation and histone modifications, which orchestrates expression of several downstream genes, for example repression of genes for ethylene production that results in superior growth traits (Ni *et al.*, 2009; Song *et al.*, 2018). Enrichment of circadian rhythm pathways has also been demonstrated in allopolyploids of *Populus simonii* (Liqin *et al.*, 2019) and of *Coffea arabica* (Bertrand *et al.*, 2015). In *Brassica rapa* paleoallopolyploids, it was demonstrated that circadian clock genes were preferentially retained during diploidization (Lou *et al.*, 2012), signifying its functional importance from an evolutionary perspective. While the alteration of circadian rhythm gene expression appears to be a typical response in allopolyploids and hybrids (Ni *et al.*, 2009; Song *et al.*, 2018), no studies have previously documented such enrichment in neoautopolyploids. We found that *LHY* gene (AT1G01060; *LATE ELONGATED HYPOCOTYL*) was differentially expressed in response to ploidy and genotype-by-ploidy interaction in the combined analysis model that included all samples. In the individual analysis models for each genotype, *LHY* was differentially expressed in response to increased ploidy in Ler but not in Sf. On the other hand, starch catabolism genes that were differentially expressed in Sf in response to increased ploidy (like *LSF1*, *LSF2*, *SEX1*, *SEX4*, *AMY3, DPE1, DPE2, FINS1, ISA3,* and *RAM1)* were not differentially expressed in Ler.

These findings reiterate the need and advantages of incorporating multiple genotypes in ploidy-based gene expression studies, as gene expression changes are likely to be genotype-specific and not passively regulated by nucleotype alone.

3.4.3 Analysing three ploidal levels enhances the ability to detect differential expression

When we compared the number of differentially expressed genes detected between diploid and tetraploid comparisons in all three analysis models (i.e. combined, individual – Ler, and individual – Sf), we found that only two genes are differentially expressed in the combined model, none in the individual analysis of Ler-only samples, and 176 in Sf-only samples (Table 3.6). When we compare these numbers to the number of differentially expressed genes obtained that respond to 'ploidy' (i.e. when the model considering all three ploidy levels - diploids, tetraploids, and octoploids together), we found 473, 149 and 1988 DEGs for each analysis model respectively. The majority of these overlap with diploid-octoploid or tetraploid-octoploid comparisons (Supplementary Data 7.13). This indicates that having an additional ploidy level as a factor in the model significantly enhances the statistical ability to detect DEGs. This could be because a third point of reference makes the gene expression patterns more evident, successfully distinguishing DEGs from background noise (transcriptome variance). The larger number of DEGs obtained between diploid-octoploid or tetraploid-octoploid comparisons also suggests that expression differences are probably scaled in these cases, making genes that would normally exhibit low-level expression differences between diploids and tetraploids (that are too small to pass the stringent thresholds of analytical tools) more prominent.

There are only a few gene expression studies that have used multiple ploidy levels to test for gene or protein expression changes. Using an autopolyploid series of haploid, diploid, triploids and tetraploids of *Zea mays,* Guo *et al.,* (1996) analysed the expression of 18 genes using northern blots and found that there was equal expression across the four ploidies on a per-genome basis for 15 of the genes and three genes showed disproportionate increase or decrease in transcript abundance relative to ploidy. Stupar *et al.,* (2007) analysed transcriptomic changes in haploids, diploids and tetraploids of *Solanum phureja* using cDNA microarrays and observed subtle but significant expression changes in 948 and 955 genes in leaflets and root tips respectively. Proteome studies in a ploidy series of inbred Oh43 revealed that 26% of the proteins were differentially expressed (Yao *et al.*, 2011). Across all the aforementioned studies, it can be seen that a significantly higher proportion of genes or proteins were detected as being differentially expressed in response to increased ploidy, which have not been detected in most other gene expression studies that compare only two ploidy levels (Albertin *et al.*, 2005; Riddle *et al.*, 2010; Allario *et al.*, 2011; Gao *et al.*, 2016; Gao *et al.*,

2016). Thus, it is advantageous to have multiple ploidy levels in the experimental design to maximise the ability to detect differentially expressed genes and understand ploidy pathways.

To our knowledge, there are no RNA-Seq studies at present that have explored transcriptomic differences across all genes for more than two ploidy levels in autopolyploids. An exception is a recent study that used a polyploidy series of *A. thaliana* to examine how ploidy affects the total RNA transcriptome size using RNA-Seq and ERCC spike-ins, however, they addressed transcriptome variation at a total transcriptome level whereas we address specific gene-level changes in gene expression (Robinson *et al.*, 2018). Previous gene-level response studies have reported findings based on a subset of all genes (e.g. 18 genes in Guo *et al.,*'s (1996) study and 9000 genes in Stupar *et al.,*'s (2007) study). It is likely that the analysis of all expressed genes would provide better insight to comprehend polyploid responses.

3.4.4 Gene expression data can be analysed in multiple ways, each providing insights on different aspects on ploidy responses

In this study, we used several different models to analyse the same dataset, each with a unique purpose to address specific questions. Two distinct analysis methods were employed. The first method was the analysis of differentially expressed genes using the statistical tool EdgeR (Robinson *et al.*, 2010; McCarthy *et al.*, 2012). Most tools to analyse differential expression tend to be very conservative as they include steps that normalize the data based on the assumption that no genes are differentially expressed. This often flattens the overall variance in the data in an attempt to account for library size differences between samples. This is a robust approach, however, genes with low-level expression differences that may be of interest are often lost as false negatives. We believe that this is the likely reason for the low number of DEGs in diploid-tetraploid comparisons. Within the DEG analysis, we analysed the data using three models. A combined model was used wherein all 24 samples were included in the analysis and the model took into account genotype effects and colchicine effects to accurately determine genes that changed in response to ploidy and genotype-by-ploidy interaction. This model gave us two important lists of genes, a list of genes that changed in response to ploidy in both genotypes, and a list of genes that changed in response to ploidy in a genotype-dependent manner. The former represents genes that have common

responses to ploidy, while the latter represents genes that responded depending on the genetic background of the plant. While this approach allowed us to determine genes that had common or genotype-specific responses, we could not isolate which of the genotype-specific responses originated from which particular genotype. To understand this, a second set of analyses were run by splitting the data by genotype. This provided insights to unique pathways that are enriched in response to increased ploidy for the two genotypes.

As the DEG analysis can be quite conservative, we also employed a second correlation-based approach to analyze the data. As there was no way to account for gene expression differences due to genotype or colchicine treatment effects for these analyses, the data from untreated diploids was excluded and the analysis was run separately for each genotype. The data was bootstrapped 10,000 times to establish thresholds for correlation coefficients and p-value based on which the genes could be declared as strongly and significantly correlated genes (SCG). This approach allowed us to identify genes that show a strong positive or negative correlation in expression values to increasing ploidy levels (Figure 3.18-Figure 3.19).

3.4.5 Detected DEGs and SCGs may be involved in regulating several pathways that determine phenotype

A number DEGs or SCGs identified in this study had known functions related to stomatal, trichome, leaf development, or flowering time traits. As quantitative traits such as these are polygenic, it remains to be tested what the contribution of each of these genes is towards the observed phenotypes. There are likely to be master genes or master regulators that respond to increased ploidy, which can have a cascading effect on a number of interacting genes and the combination of their expression determine the final phenotype. *GIGANTEA* could potentially be one such master gene in ploidy responses as it is a key gene that plays a master regulatory role in several pathways like drought tolerance, salt tolerance, circadian clock control, starch accumulation, light signaling, chlorophyll accumulation, and cold tolerance (Mishra and Panigrahi 2015). We found that *GIGANTEA (GI)* was differentially expressed in response to ploidy as well as genotype-by-ploidy interaction. If we examine the enriched GO terms and KEGG pathways (Table 3.20-3.21), we find that almost all pathways (except drought tolerance) that *GIGANTEA* is known to be involved in are enriched. In diploidization following

paleopolyploidy, it has been shown that circadian clock genes are preferentially retained over time (Lou *et al.*, 2012). It is interesting that at least two large-effect circadian-clock genes (*GIGANTEA* and *LHY*) are differentially expressed in the early stages after neoautopolyploid formation, suggesting that their retention may not entirely be determined by evolutionary processes over time but may in fact have other processes regulating them almost immediately after ploidy increase. Circadian clock genes are thought to regulate a third of the expressed genes in *A. thaliana* and are typically enriched in stress response pathways and plant hormone pathways (Covington *et al.*, 2008).

Chapter 4 : **NATURAL GENETIC VARIATION IN RESPONSE TO INTERPLOIDY HYBRIDIZATION IN *ARABIDOPSIS THALIANA***

Sarah Chordekar and Paula X. Kover

Author Contributions:

SC: performed the experiments, analysed the data, and wrote the draft

PXK: supervised, performed the experiments, analysed the data, and reviewed the draft.

4.1 INTRODUCTION

It has been suggested that changes in ploidy may cause significant barriers to genetic exchanges with relatives of a different ploidy, contributing to the reproductive isolation and fast speciation of newly formed polyploids (Stebbins Jr., 1950; Grant, 1981; Sweigart *et al.*, 2008; Köhler *et al.*, 2010). Successful interploidy hybridization requires the bypassing of several obstacles, either in the form of pre-zygotic hybridization barriers i.e. barriers before fertilisation like pollen incompatibility, or post-zygotic hybridization barriers i.e. barriers after fertilisation like hybrid sterility (Köhler *et al.*, 2010). Reduction in interploidy cross fertility is commonly observed in triploid F1 progeny that result from crosses between tetraploid and diploid plant species. This is termed as a "triploid block" because triploid seeds often fail to develop (Marks, 1966; Johnston & Hanneman, 1982) due to the formation a post-zygotic hybridization barrier in the endosperm which leads to seed abortion (Erilova *et al.*, 2009). However, under some circumstances, triploid seeds can be viable and act as bridges allowing gene flow between ploidy levels when backcrossed to their progenitors (Ramsey & Schemske, 1998; Husband, 2004). Fertile triploid seeds produce triploid plants that are most often transient due to meiotic instability owing to the odd number of chromosome pairs. The following generation, thus, usually consists of a range of mixed ploidy progenies (Henry *et al.*, 2005) that are either unstable and perish, or are stable and lead to new species lineages (Husband, 2004; Vallejo-Marín & Hiscock, 2016; Kopecký *et al.*, 2018). Thus, understanding the mechanisms underlying the success or failure of interploidy crosses would be useful for understanding the role of polyploidy in plant speciation.

The mechanisms identified to be involved in inter-ploidy hybridization failure are thought to be similar to those that underlie interspecies hybridization barriers as both involve altered endosperm cellularisation and embryo lethality (Sukno *et al.*, 1999; Bushell *et al.*, 2003; Schatlowski & Kohler, 2012; Lafon-Placette *et al.*, 2018). The genes identified to be involved in interploidy hybridization response like *Polycomb Repressive Complex2* (*PRC2*) subunits *MEDEA (MEA)* and *FERTILIZATION INDEPENDENT SEED 2 (FIS2)*, and *AGAMOUS LIKE (AGL)* genes have also been identified to play a role in interspecies hybridization response (Josefsson *et al.*, 2006; Erilova *et al.*, 2009; Rebernig *et al.*, 2015). Hybridization between crop varieties and wild relatives is a potentially

valuable source of genes for crop improvement, but inviability of their hybrids often hamper the success of this approach (Hajjar & Hodgkin, 2007; Dempewolf *et al.*, 2017; Prohens *et al.*, 2017). Thus, understanding the genetic basis of variation in interploidy cross success can also be helpful to understand how interspecies barriers might be broken.

In most angiosperms, failure in crosses between plants of different ploidies is attributed to endosperm development failure, leading to seed abortion of varying degrees (Ramsey & Schemske, 1998; Scott *et al.*, 1998; Adams *et al.*, 2000; Dilkes *et al.*, 2008; Erilova *et al.*, 2009; Ishikawa *et al.*, 2011; Stoute *et al.*, 2012; Hehenberger *et al.*, 2012; Schatlowski & Kohler, 2012; Sekine *et al.*, 2013; Lafon-Placette & Köhler, 2016; Schinkel *et al.*, 2017). The endosperm, unlike other plant tissues, contains an unbalanced set of chromosomes: two copies of the maternal genome ("2m" from the homodiploid maternal central cell) and one copy of the paternal genome ("1p" from the haploid paternal sperm cell). This specific 2m:1p ratio called the endosperm balance number (EBN) is crucial for normal seed nourishment and development. A departure from this balance has been shown to either accelerate (when there is more than 2:1 maternal excess) or delay endosperm cellularization (when there is more than one paternal copy for every two maternal copies) (Lin, 1984; Scott *et al.*, 1998; Sorensen *et al.*, 2002; Costa *et al.*, 2004; Dilkes & Comai, 2004; Pignocchi *et al.*, 2009). Several regulators of endosperm cellularisation like *AGAMOUS LIKE (AGL)* transcription factor *AGL62* (Erilova *et al.*, 2009; Lu *et al.*, 2012) and *TRANSPARENT TESTA GLABRA2 (TTG2)* (Dilkes *et al.*, 2008) have been identified to be directly associated with promoting seed abortion by delaying endosperm cellularisation in paternal excess interploidy crosses. Parental-dosage sensitivity has been proposed to be due to de-regulation of genomic imprinting patterns in the endosperm (Haig & Westoby, 1991; Gutierrez-Marcos *et al.*, 2003; Jullien & Berger, 2010; Schatlowski & Kohler, 2012). This hypothesis is supported by the fact that interploidy crosses can be rescued by altering imprinting patterns, for example, hypomethylated pollen derived from the *met1* mutant could bypass the interploidy hybridization barrier in *Arabidopsis thaliana* (Schatlowski *et al.*, 2014b). If parental-dosage imbalance is the main trigger for seed abortion in response to interploidy crosses, seed abortion should occur in a consistent proportion between parents of different ploidies, independent of the specific maternal and/or paternal

genotypes involved in the interploidy cross. Despite the extensive research on understanding parent-of-origin effects in interploidy crosses, very little is known about maternal and paternal genotypic variation in interploidy cross success and its genetic basis (but see Henry *et al.*, 2005; Dilkes *et al.*, 2008).

Interspecies crosses also show parent-of-origin effects, with similar directional effects on endosperm cellularization as seen in interploidy crosses (Valentine & Woodell, 1963; Ishikawa *et al.*, 2011; Rebernig *et al.*, 2015; Garner *et al.*, 2016). This suggests that dosage sensitivity issues in endosperm development may arise independent of ploidy differences. It is possible that interspecific crosses cause imprinting deregulation, similar to what is observed in interploidy crosses (Chen *et al.*, 2016; Florez-Rueda *et al.*, 2016). However, there is also evidence that despite homoploidy between some species (i.e. two species may have the same number of complementary genomes), they may have different "effective ploidy" (Johnston & Hanneman, 1982; Lafon-Placette & Köhler, 2016; Roth *et al.*, 2019). This suggests that species may experience significant gene expression divergence that cause dosage imbalances during endosperm development (Roth *et al.*, 2019). Imbalances may also occur due to interactions between maternal and paternal genomes in the endosperm. Kirkbride *et al.*, (2015) crossed three different *A. thaliana* genotypes to *A. arenosa* (a tetraploid sister-species) and found that maternally expressed genes were generally conserved in all F1 hybrids but paternally expressed genes were differentially expressed depending on the maternal genotype contributing to the hybrid. Thus, this provides evidence for an interaction between maternal and paternal genotypes in the developing seed which may be involved in determining the extent of incompatibility or seed abortion.

Our study focusses on understanding if the seed abortion response in interploidy hybridization is genotype-dependent and whether the maternal and paternal genotypes independently regulate the outcome of an interploidy cross or whether it is determined by the interaction between the two. The limited data currently available on this provides opportunities to further explore this avenue and further understand factors responsible for post zygotic hybridization barriers in interploidy crosses which can have implication on our understanding of interspecies hybridization barriers as well as plant

speciation. We use *A. thaliana* as a model as it provides an interesting system to investigate these issues because unlike many plants, they are tolerant to interploidy crosses, producing triploid seeds with variable success. Due to this fact, *A. thaliana* has been used in many studies to investigate molecular mechanisms involved in endosperm failure, or the success of triploid progeny (Scott *et al.*, 1998; Henry *et al.*, 2005; Dilkes *et al.*, 2008; Erilova *et al.*, 2009; Schatlowski & Kohler, 2012; Lafon-Placette & Köhler, 2016). However, despite the wide usage of natural variation in *A. thaliana* to study the genetic basis of many quantitative traits (Kover *et al.*, 2009; Camargo *et al.*, 2018; Ma *et al.*, 2019; Yang *et al.*, 2019) there is limited information about the extent of natural genetic variation in *A. thalina*'s tolerance to interploidy crosses. Nevertheless, Dilkes *et al.,* (2008) performed interploidy crosses using two different genotypes of *A. thaliana* and demonstrated the potential for quantitative variation in the success of interploidy cross among different natural genotypes of this species.

In order to identify natural allelic variation underlying a quantitative trait, synthetic populations of recombinant inbred lines (RILs) derived from two-parent crosses are commonly used. Here, we use *A. thaliana* MAGIC (Multi-Parent Advanced Genetic Inter-Cross) lines, a type of RIL derived from a multiparent population of 19 founder genotypes that have been inter-crossed for four generation and inbred for a further seven generations (Kover *et al.*, 2009). All lines have been sequenced for 1260 Single Nucleotide Polymorphisms (SNP's), which allow high-resolution quantitative trait loci (QTL) mapping. The lines are stable and do not need to be re-genotyped for each experiment. Using a multiparent population provides several advantages over traditionally used two-parent populations as they facilitate higher resolution mapping resulting in a smaller number of candidate genes, reduce the confounding effects of linked loci, and produce information-rich patterns of effects that can help identify causal variants and distinguish pleiotropy from chance colocalization of multiple QTL (Broman *et al.*, 2019).

Our experiment was conducted in two stages – a pilot stage to establish proof of concept and determine if there is any natural variation in the rate of seed abortion in interploidy crosses; and the main experiment stage where the genetic variation was fine mapped. In the first pilot experiment, we assessed natural genetic variation in tolerance

to interploidy crosses by evaluating rates of seed abortion in response to paternal excess [2X(♀) x 4X(♂)] crosses among seven naturally occurring genotypes of *A. thaliana*, to address the following questions: (1) Does seed abortion vary in response to the maternal genotype. (2) Does seed abortion vary in response to the paternal genotype. (3) Is the combination of the maternal and paternal genotype important in determining the response? Based on our findings from the pilot study, we found results that would benefit from being scaled up to include a larger number of genotypes. For this, in the second stage of the experiment, we used MAGIC lines to map genetic factors involved in the genetic variance in response to paternal excess interploidy crosses, and to determine if the genetic basis of this response to different paternal genotypes were independent - a question that has not been addressed to date.

Loci identified as having a significant QTL for interploidy seed abortion response were scanned for candidate genes in a ± 250kb region around the QTL marker and were shortlisted if they were involved in one or more of the following functions : (1) Auxin synthesis, binding or response : Auxin is a key plant hormone involved in growth and development of the endosperm following fertilization by triggering cell division of the central cell (Figueiredo *et al.*, 2015, 2016); (2) Pollen development : when a pollen grain lands on the stigma of a maternal plant, it undergoes hydration and the pollen tube germinates and penetrates into the style making its way to the ovary. During the journey of pollen cells, they interact with at least seven types of cells in the maternal reproductive organs (Palanivelu & Tsukamoto, 2012). Thus, pollen tube growth and delivery of pollen cells to the egg cell is partly regulated by the maternal plant and the interaction of the two genotypes may impact the fertility of the interploidy cross; (3) Cytoskeletal regulation and cell division : since interploidy crosses cause changes in the overall genomic content of all cells involved, it poses a challenge to sister chromatid pairing during mitosis, regulation of DNA replication and cell division; (4) Development and differentiation of the embryo or endosperm. We propose that genotypic variation for one or more of these genes could be linked to a host of post zygotic changes, consequently leading to variation in seed abortion based on the maternal or paternal genotypes contributing to the cross.

4.2 MATERIALS AND METHODS

4.2.1 Plant material

For the pilot study, to scan for natural genetic variation in seed abortion in response to interploidy hybridization in *A. thaliana,* we used seven diploid genotypes as the maternal parents (Bur, Col, Kn, Ler, Mt, Sf, Wil). These seven genotypes are part of the 19 founder genotypes of the MAGIC lines (Multi-Parent Advanced Genetic Inter-Cross) (Kover *et al.*, 2009) that have been used as maternal genotypes for interploidy crosses in the main mapping experiment. Seeds for the pilot study were initially obtained from the Arabidopsis Biological Resource Centre (ABRC, USA) and maintained for several generations in the lab. Neo-tetraploid plants were generated, tested for ploidy and grown for three generations as described previously in Chapter 2.

For the mapping experiment we used a random selection of 261 *A. thaliana* MAGIC lines as the diploid maternal parents (Supplementary Data 7.14). The complete MAGIC population consists of over 700 lines that are the result of four generations of intercrossing among 19 genotypes, followed by six generations of inbreeding. Multi parent mapping lines offer higher resolution that two parent populations (Broman *et al.*, 2019). These lines are near-homozygous and stable and thus do not need re-genotyped for each experiment. Furthermore, the lines have been genotyped with 1,260 single nucleotide polymorphisms that allow fine mapping of quantitative trait loci (QTL) to a mapping accuracy of ± 250kb (Kover *et al.*, 2009). For this study, we use a randomly selected subset of the MAGIC population and grew three to six replicates for each MAGIC line to be crossed to pollen from diploid lines, neo-tetraploid lines and itself (described below).

4.2.2 Growth Conditions

Seeds were sown in 2” pots containing soil (F2+S variety of compost from Levington® Seed and Modular Compost, Scotts Company, UK) and randomly allocated to trays that were rotated regularly to avoid any positional effects. To ensure that there were limited number of plants to cross at any given point, the plants were grown in batches either in the growth chamber or glasshouse set at 21°C day/18°C night and 16 hours of light/day.

4.2.3 Plant Crosses

For all crosses, early flower buds (between 5th to 15th bud on the primary stem) from plants chosen as the maternal parents were emasculated prior to anthesis and any open flowers were removed to avoid cross-contamination with self-pollen. Stigmas of emasculated buds were allowed to mature for one day and then hand-pollinated with pollen from the required paternal plants. The pedicel of the crossed bud was tagged using a coloured thread to identify the paternal plant it had been crossed to. Each maternal individual produced a single cross to each paternal genotype. Thus, biological replicates of each cross type are derived from independent maternal plants. Three to five biological replicates were produced of each cross pair.

For the pilot experiment, crosses were conducted between diploid maternal plants and tetraploid paternal plants within and between genotypes with an aim to create a 7x7 matrix of interploidy, inter-genotype crosses with at least three biological replicates. However, due to differences in flowering time between some genotypes, we could not achieve all 49 combinations. Nevertheless, the pilot study provided sufficient information to pursue and design the main experiment.

For the main experiment, maternal plants for each MAGIC line were crossed to five pollen parents - Col 2x (C2), Col 4x (C4), Mt 2x (M2), Mt 4x (M4) and self-cross (an open flower from the maternal plant was used as the pollen parent and hand-pollinated). Three to five biological replicates from independent maternal plants were produced for each cross pair. The C2 and M2 crosses served as genotype controls to account for any seed abortion that may be caused due to interaction between genotypes of the parents contributing to the crosses; the self-cross (S) served as an indicator of maternal plant health to ensure that any abortion effects seen are on account of genotype and/or ploidy only and not due to the plant being stressed or unhealthy. Crosses were allowed to develop into siliques that were grown to maturity and harvested for further analysis approximately three weeks after crossing.

As the MAGIC lines are a mosaic of 19 contributing founder genotypes, it was important to establish whether the abortion response we see is solely due to interploidy hybridization or if it is also due to any genetic incompatibility between the MAGIC lines and the paternal genotypes (Col and Mt). Thus, in addition to the MAGIC line x 4x

tetraploid crosses for the two paternal genotypes (C4 and M4), we also conducted corresponding diploid crosses (MAGIC lines x 2x) for both paternal genotypes (C2 and M2) as well as crosses to self-pollen (S) which was a technical control.

A total of 229 and 2,215 mature siliques were collected for the pilot study and main experiment respectively.

4.2.4 Seed Imaging and scoring

Harvested seeds from crossed siliques were imaged under 14x magnification using a macro lens (Olloclip®, USA) coupled to an iPhone camera (Apple, USA). The seeds were scored into two categories - aborted and non-aborted. Aborted seeds were characterised as small, shriveled and dark brown or black coloured seeds (Figure 4.1a). The remainder of the seeds were either abnormal (Figure 4.1b), characterised by irregular morphology and light brown coloured seeds, or normal (Figure 4.1c), characterised by plump, elliptical, light brown coloured seeds. Both abnormal and normal seeds were categorised as non-aborted (Figure 4.1b-c). Due to unequal number of seeds produced in each silique, we analyzed the percentage of aborted seeds (calculated by dividing the number of aborted seeds by the number of total seeds produced in each cross and multiplying that by 100). Only cross pairs that produced a minimum on ten seeds were used for the analysis.

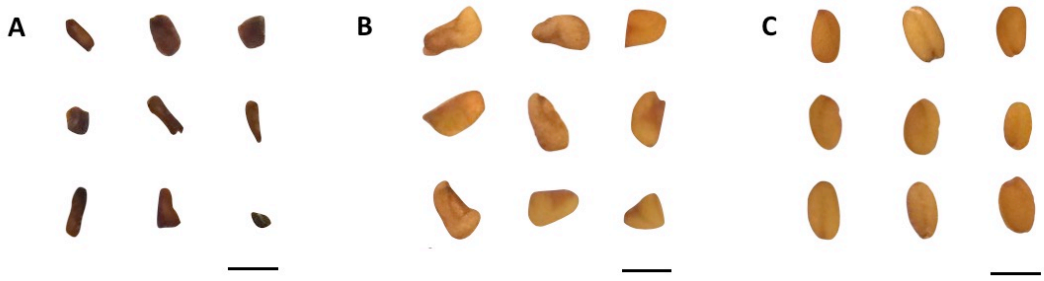


Figure 4.1 | Seed types observed among triploid seeds produced by crossing diploid maternal plants and tetraploid paternal plants. Shrivelled seeds – categorized as aborted; (B) Abnormal seeds - categorized as non-aborted; (C) Normal seeds - categorized as non-aborted. Scale bar = 500µm.

4.2.5 Statistical Analysis

4.2.5.1 Pilot Experiment

To determine whether there was significant genetic variation among maternal and paternal genotypes, as well as a significant interaction between maternal and paternal genotypes on the percentage of seeds aborted in the pilot experiment, we used a two-way ANOVA.

4.2.5.2 Main Experiment

4.2.5.2.1 Determining genotype incompatibilities in C2 and M2 relative to the self-cross (S)

The effect of paternal line on percentage seed abortion in the main experiment was assessed using a one-way ANOVA followed by a Tukey's Post-hoc test to compare C2 and M2 crosses to S to determine genotype incompatibility between the MAGIC lines and the paternal genotypes C2 and M2. This was tested to ascertain that any responses we saw in inter-ploidy seed abortion were due to the effects of ploidy and genotype-by-ploidy interaction and not due to incompatibilities between the MAGIC lines and paternal genotypes being used in the study.

4.2.5.2.2 Determining the effect of maternal and paternal genotypes in MAGIC x 4x crosses

To determine whether there was significant genetic variation among maternal and paternal genotypes, as well as a significant interaction between maternal and paternal genotypes on the percentage of seeds aborted in crosses between the MAGIC lines and the paternal tetraploid genotypes, we conducted a two-way ANOVA.

4.2.5.2.3 Regression analysis between MAGIC x C4 and MAGIC x M4 crosses

The percentage seed abortion data for MAGIC x C4 and MAGIC x M4 crosses was used to perform a linear regression analyses to determine if there was any relationship between seed abortion in crosses to the two paternal tetraploid genotypes. The coefficient of determination (R^2) was calculated to evaluate the fit of data points around the regression line and estimate what proportion of the variance in MAGIC x M4 that could be explained by variance in MAGIC x C4 crosses. If there was no natural variation in response to interploidy crosses with different paternal genotypes, the slope of the regression line would be expected to be unity (m=1) and the intercept zero (c=0). To determine the extent to which the regression line deviates from unity, the 'smatr' R

package function 'sma' was used as follows: s*ma(y ~ x -1, log="xy", data=data, slope.test=1).* This function tests if the slope of the regression line passing through origin is significantly different from the null hypothesis of slope=1.

All statistical analyses were done in R (R Core Team, 2018).

4.2.6 QTL Mapping

To uncover genetic factors underlying the variation in seed abortion in response to interploidy crosses, we performed a QTL analysis on the percentage of seeds aborted when the paternal parent was a C4 and a M4 independently. The best estimate for the %abortion (percentage of aborted seeds) phenotype for each MAGIC line was estimated as the average across replicates of the cross. QTL mapping was performed in R using the 'HAPPY' package as described in (Kover *et al.*, 2009). Briefly, a probabilistic reconstruction of the genome of each MAGIC line as a mosaic of the 19 contributing parental haplotypes is calculated taking into account information from the 1260 Single Nucleotide Polymorphisms (SNPs) using a hidden Markov model. Next, the entire genome is scanned for evidence of a QTL in each SNP interval using a fixed effects linear model ANOVA assuming that there is no other QTL. The statistical significance of the genome scan at each SNP interval is calculated as logP= $-\log_{10}$(ANOVA P-value). Thus, the higher the significance of a QTL, the greater its logP value. The evidence of each QTL is then re-evaluated in the context of a multiple QTL mapping model. A QTL is considered significant when logP>3 and genome-wide P<0.1. It is expected that, if interploidy cross failure is due to genetic variation in maternal plant's ability to detect ploidy, then the two independent QTL analyses for each tetraploid paternal genotype will uncover similar QTL profiles.

4.2.7 Shortlisting candidate genes

After QTL locations for %abortion were identified in crosses with the paternal parents - C4 and M4, a list of genes in a ± 250kb region around the QTL marker chromosomal locations were obtained using the Arabidopsis annotation (Araport11_GFF3_genes_transposons.201606.gtf) downloaded from Araport (www.araport.org; accessed 23/03/2018). Gene descriptions and annotations for these genes were looked up on TAIR (https://www.arabidopsis.org; accessed 27/11/2018) and

protein coding genes with known functions were retained. Genes considered as good candidates as the causative genetic factor for each QTL location were ones with known function involving: (i) auxin production and regulation; (ii) pollen development; (iii) cytoskeletal organization; (iv) mitosis, cell cycle or cell division; (v) seed development and (vi) cell differentiation.

4.3 RESULTS

4.3.1 Pilot Study – establishing natural variation among Arabidopsis genotypes in interploidy cross success

When diploid maternal plants from seven different genotypes were crossed to tetraploid fathers, we saw significant effects of the maternal genotype and paternal genotype on the percentage of seeds aborted. No significant maternal-by-paternal interaction effects were seen (Table 4.1). Figure 4.2 shows a representative plot of these effects using a subset of the data. A subset of crosses with two paternal tetraploids (Col-4x and Ler-4x) is chosen due to missing cross pairs of maternal genotypes with some of the other paternal genotypes. We observed that when Col-4x pollen were crossed to various maternal genotypes, the least percent seed abortion was observed when crossed to Mt and Sf mothers, and the most percent seed abortion was observed when crossed to Wil, Col and Kn mothers. On the other hand, when Ler-4x pollen were crossed to various maternal genotypes, the least percent seed abortion was observed when crossed to Col and Ler mothers, and the most percent seed abortion was observed when crossed to Wil and Kn mothers. Furthermore, all maternal plants crossed to Ler-4x pollen showed a lower percentage of aborted seeds in comparison to crosses with Col 4x pollen for the same maternal genotypes. Thus, the percentage of seed abortion in the paternal excess interploidy crosses varied depending on the maternal genotypes, paternal genotypes and the combination of genotypes contributing to the cross.

Table 4.1 | ANOVA output for maternal and paternal effects in interploidy crosses – pilot experiment. Two-way ANOVA results for the effect of maternal genotype, paternal genotype and maternal-by-paternal interaction.

Factor	Df	Sum Sq	F value	P-value
Paternal Genotype	6	32423.680	8.511	1.51E-06
Maternal Genotype	5	26203.555	8.254	7.19E-06
Paternal x Maternal	10	10737.674	1.691	1.06E-01

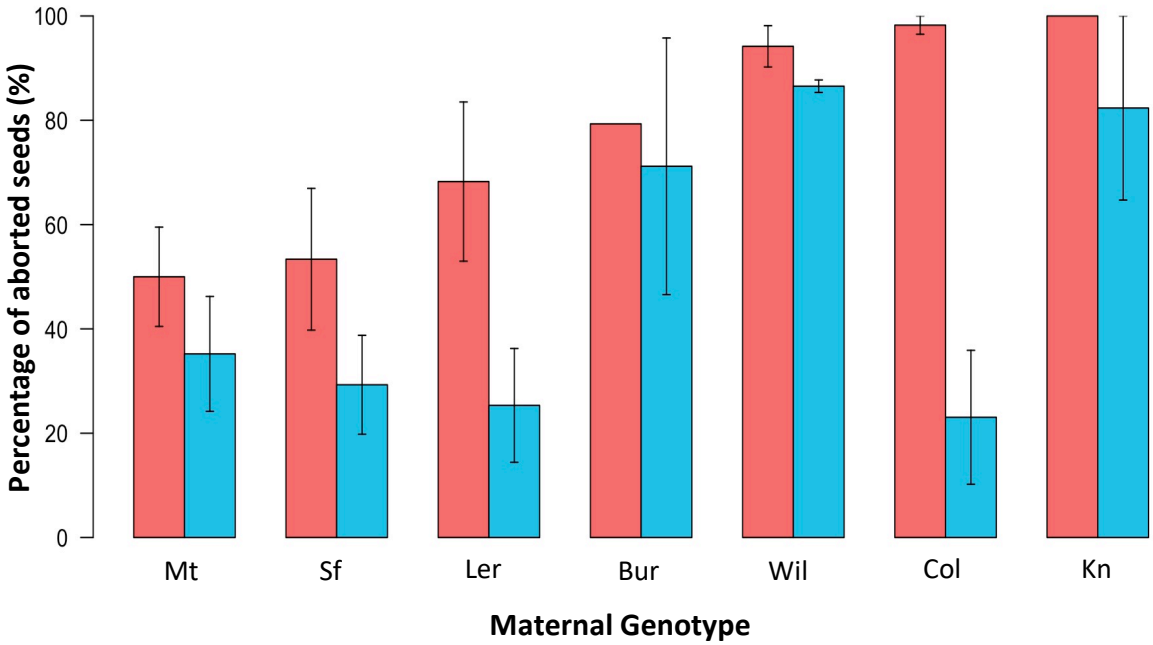


Figure 4.2 | Natural variation in response to interploidy crosses. Maternal and paternal variation in percentage seed abortion in seven maternal genotypes crossed to two paternal genotypes. Bars indicate percentage seed abortion averaged across the biological replicates. Red bars indicate the maternal genotype (x-axis) crossed to Col-4x pollen. Blue bars indicate the maternal genotype (x-axis) crossed to Ler-4x pollen. Error bars represent ± 1 SE (standard error of the mean).

4.3.2 Seed abortion is a response to interploidy crosses

Figure 4.3 illustrates that across all MAGIC maternal genotypes, the percentage of seed abortion in crosses to C2 or M2 paternal genotypes is nearly 0% barring a few outliers, and interploidy crosses to C4 or M4 paternal genotypes show a wide range of percentages of seeds aborted. To establish whether the abortion response we saw was solely due to interploidy hybridization or if it was also due to any genetic incompatibility between the MAGIC lines and the paternal genotypes (Col and Mt), we performed a one-way ANOVA on the percentage of seeds aborted as function on paternal type (C2, M2, M4, C4 and S) and found a significant paternal type effect (Table 4.2). However, a subsequent Tukey's post-hoc test showed that there was no significant difference between the percentage of seeds aborted between crosses with C2 and S pollen, and between crosses with M2 and S pollen (Table 4.3), and the main effect is cause by a significantly higher percentage of aborted seeds in crosses with tetraploid pollen (Table 4.3). This suggested that seed abortion was mainly a response to interploidy crosses,

and that the genetic background of the tetraploid pollen would not result in any seed abortion due to additional genetic incompatibility between the maternal and paternal genotypes *per se*. Furthermore, the Tukey's test also showed that crosses to a M4 pollen resulted in significantly lesser abortion when compared to C4 pollen, thus depicting a clear paternal genotype effect in interploidy crosses, independent of ploidy.

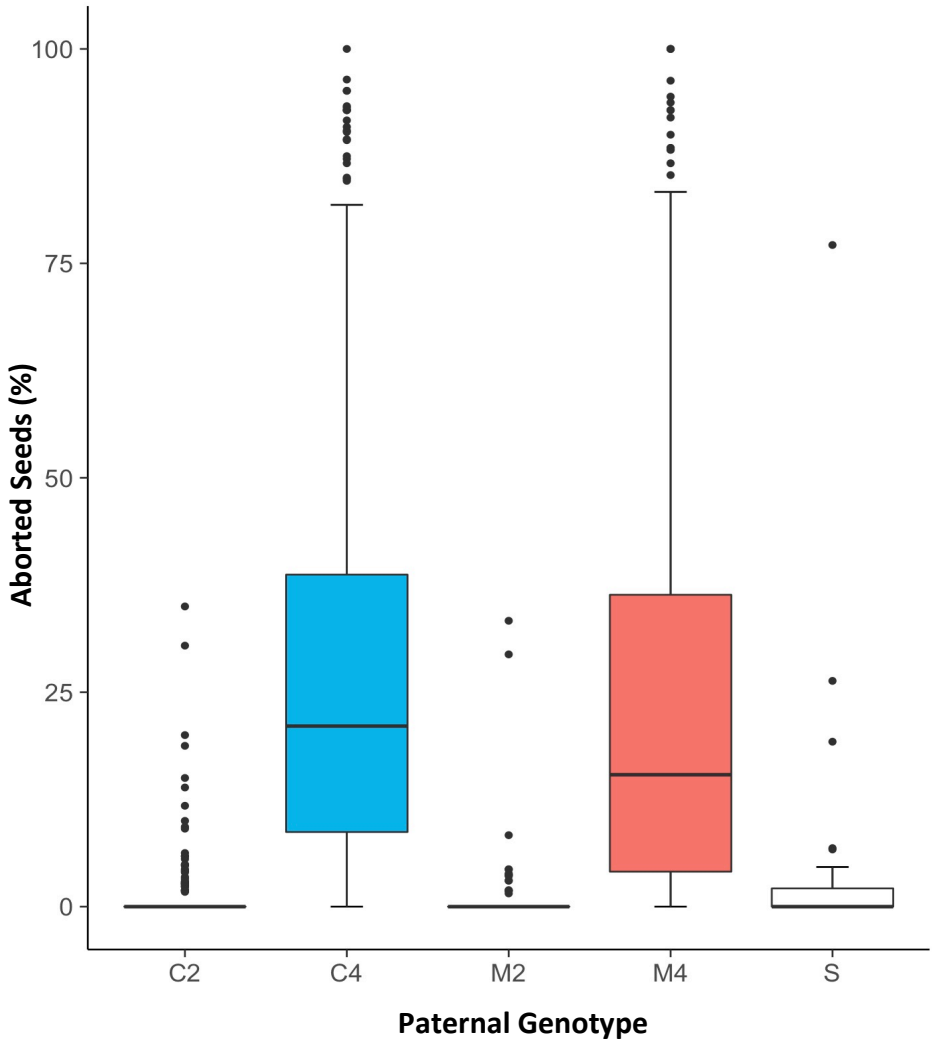


Figure 4.3 | Boxplot of percent seed abortion in the MAGIC lines crossed to five pollen parents. 261 MAGIC lines (used as maternal plants) were each crossed to pollen from C2 (Col-2x), C4 (Col-4x), M2 (Mt-2x), M4 (Mt-4x) and S (self) paternal plants. For each data point, the average percent seed abortion from three to five biological replicates for each cross pair was used. Horizontal lines represent the median percent seed abortion, boxes represent the interquartile range, whiskers indicate the minimum and maximum values and the black dots represent outliers.

Table 4.2 | One-way ANOVA output for paternal effects in MAGIC lines crossed to various paternal genotypes.

Factor	Df	Sum Sq	F value	P-value
Paternal Genotype	4	185239.940	107.363	1.85E-80

Table 4.3 | Tukey's HSD post-hoc test output comparing crosses between MAGIC lines with different paternal genotypes. Lower and upper bound represent the lower and upper limits of the 95% confidence interval. Sig represents the adjusted P-value for the comparison corrected for multiple comparisons.

Contrast	Mean Diff	Lower Bound	Upper Bound	Sig
S-C2	1.472	-6.704	9.649	9.88E-01
S-M2	2.231	-6.024	10.486	9.48E-01
S-C4	-24.777	-32.296	-17.258	**0.00E+00**
S-M4	-20.425	-27.919	-12.930	**0.00E+00**
M2-C2	-0.759	-6.629	5.111	9.97E-01
M2-C4	-27.008	-31.922	-22.094	**0.00E+00**
M4-C2	21.897	17.156	26.638	**0.00E+00**
M4-C4	-4.352	-7.840	-0.865	**6.04E-03**
M4-M2	22.656	17.781	27.531	**0.00E+00**
C4-C2	26.249	21.468	31.030	**0.00E+00**

4.3.3 The seed abortion responses to different paternal genotypes were significantly different

A two-way ANOVA testing for the effects of maternal and paternal genotype on percentage seed abortion in the interploidy crosses revealed that significant effects were seen for maternal genotype as well as paternal genotypes, however, no significant maternal-by-paternal interaction effects were observed (Table 4.4). Thus, the genotypes of the maternal and paternal plant independently affected seed abortion and their specific combination did not appear to have an effect.

Table 4.4 | ANOVA output for maternal and paternal effects in interploidy crosses – main experiment. Two-way ANOVA results for the effect of maternal genotype, paternal genotype and maternal-by-paternal interaction in MAGIC diploids crossed to two tetraploid genotypes.

Factor	Df	Sum Sq	F value	P-value
Paternal Genotype	1	5010.056	13.311	2.87E-04
Maternal Genotype	256	322736.399	3.350	1.34E-33
Paternal x Maternal	215	87698.597	1.084	2.31E-01

Figure 4.4 depicts the scatterplot of average percentage seed abortion in MAGIC x M4 crosses v/s MAGIC x C4 crosses for 261 MAGIC lines. The regression line (blue) is significantly different from the theoretical 1:1 reference line (black; slope=1) suggesting that percent abortion in MAGIC lines crossed to C4 pollen is overall significantly different to that of the same MAGIC line crossed to M4 pollen. A correlation test on the average percentage seed abortion data for MAGIC x M4 crosses v/s MAGIC x C4 crosses suggested that there was a moderate correlation between the two (Pearson's r=0.531) and that 28% of the variability in MAGIC X C4 crosses could be explained by the linear

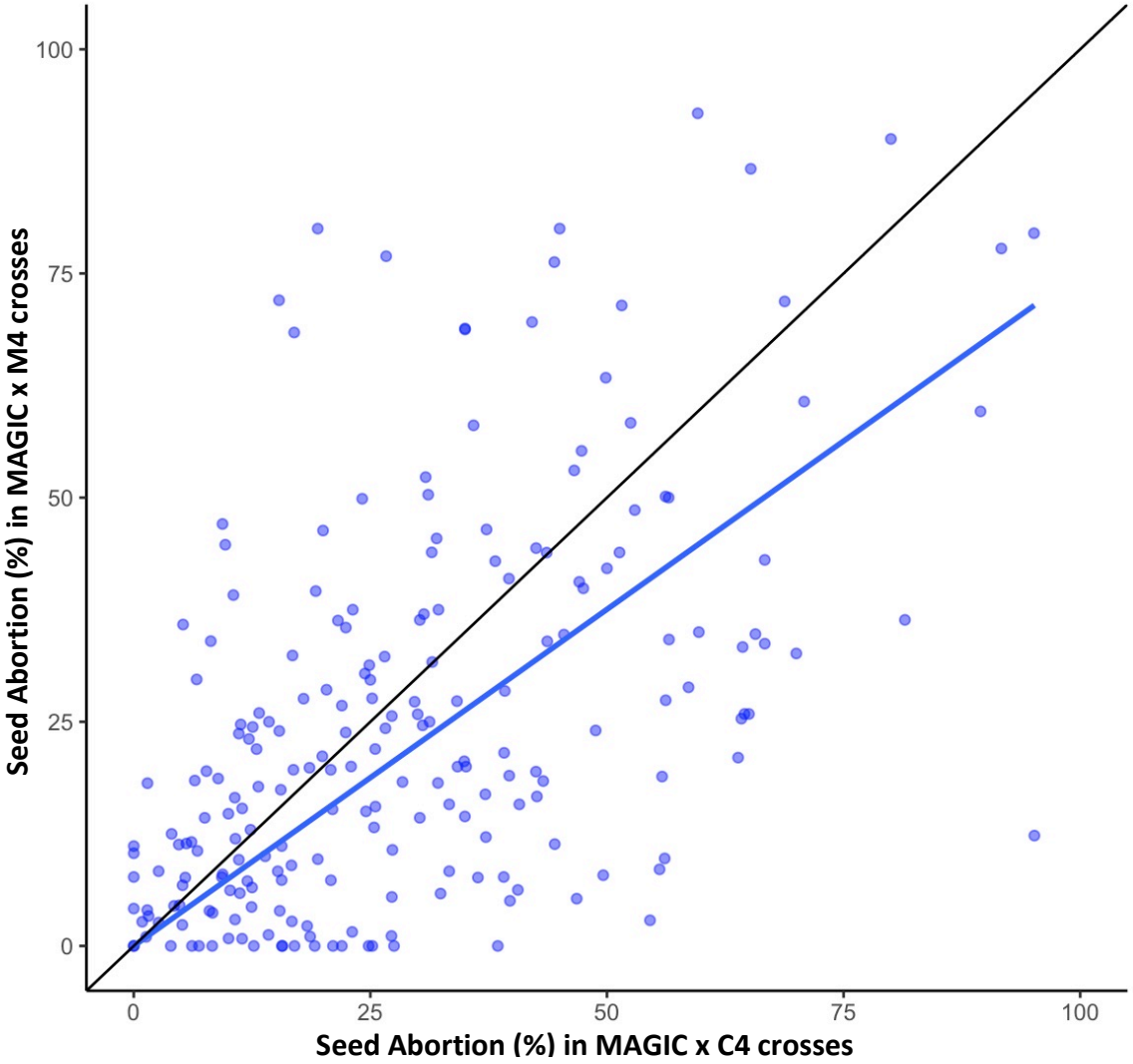


Figure 4.4 | Scatterplot of percent seed abortion in MAGIC x C4 v/s MAGIC x M4 crosses. Blue dots indicate percent seed abortion for 261 MAGIC lines averaged across three to five replicates crossed to C4 (y-axis) or M4 (x-axis) pollen. The blue line indicates the linear regression line (model=y~x-1); Pearson's correlation coefficient (r) =-0.139; p-value=0.0398; slope=0.904. The black line indicates the reference line of slope=1 passing through origin. N=1059.

model (R^2=0.28). Thus, there was little similarity in seed abortion response to C4 and M4 pollen when crossed to the same maternal MAGIC lines and most of the variation in percentage seed abortion (72%) remained unexplained.

4.3.4 The genetic basis of triploid block responses to different paternal genotypes were independent - QTL Mapping for seed abortion in MAGIC lines crossed to C4 and M4

To further understand the genetic architecture behind the variation in interploidy crosses between the MAGIC lines and the two paternal genotypes, we performed a QTL mapping analysis for percentage seed abortion and identified one QTL on chromosomes 3 when seeds had C4 paternal parents, and 1 QTL on chromosome 4 when seeds had M4 paternal parents (Table 4.5; Figure 4.5). The QTL for seed abortion with C4 and M4 were located on separate chromosomes suggesting that the genetic basis of maternal response to crosses with tetraploid pollen differed depending on the paternal genotype.

Table 4.5 | Significant QTL detected for average percent abortion for MAGIC line crosses with C4 and M4 paternal plants. QTLs where the logP of genetic association is genome-wide significant with a permutation P-value<0.1 are listed below. "Chromosome" indicates the chromosome number the QTL is located on, "peak (kb)" the position of the QTL in the chromosome in kilobases.

Trait	Chromosome	peak (kb)	logP	genome-wide P
MAGIC x C4	3	10,141	3.189	0.095
	3	10,381	3.190	0.095
	3	12,489	4.176	0.009
MAGIC x M4	4	1,239	5.466	0.002

The estimated effect of each of the 19 parental genotypes of the MAGIC lines on maternal response to interploidy cross (Table 4.6) showed that the maternal alleles contributing to the highest seed abortion in crosses with C4 pollen were alleles from Po, Oy and Rsch genotype. In contrast, the maternal allele that contributed to the highest abortion in crosses with M4 pollen was the allele from the Ws genotype. This further reiterated that most of the genetic basis for the response to interploidy crosses was independent and determined by specific alleles of the maternal and paternal genotypes.

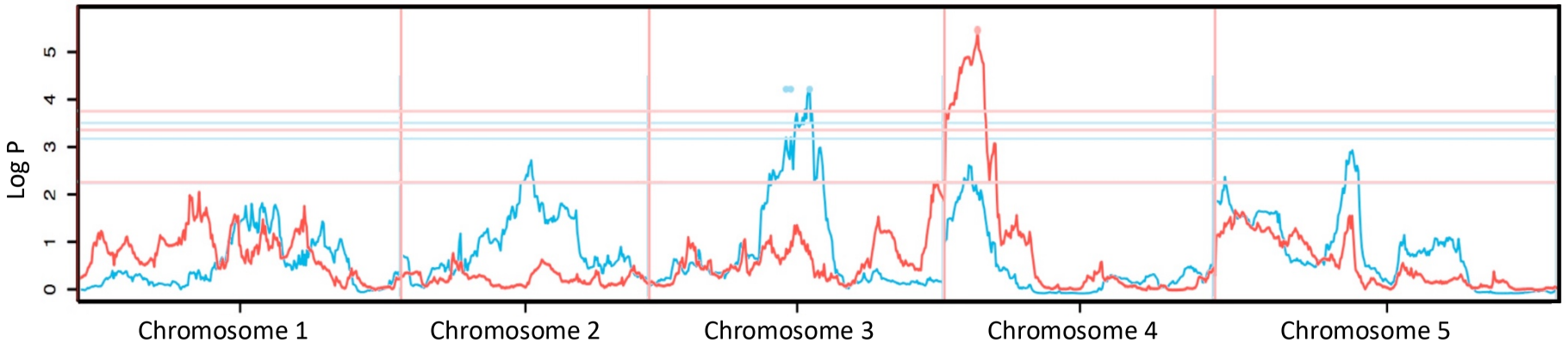


Figure 4.5 | Superimposed QTL scans for percent seed abortion in MAGIC x C4 crosses (blue) and MAGIC x M4 crosses (red).

Significant QTLs are marked with blue and red dots for C4 and M4 crosses respectively. Genomewide p<0.1

Table 4.6 | Estimated seed abortion value for each of the 19 founder genotypes of the MAGIC lines at the QTL locations.

Chr	Mb	Bur-0	Can-0	Col-0	Ct-1	Edi-0	Hi-0	Kn-0	Ler-0	Mt-0	No-0	Oy-0	Po-0	Rsch-0	Sf-2	Tsu-0	Wil-2	Ws-0	Wu-0	Zu-0
MAGIC X C4																				
3	10.1	35.2	46.4	37.7	43.3	47.8	35.1	40.3	30.2	42.7	27.2	54.8	57.7	48.2	32.6	52.1	41.6	39.8	37.1	22.9
3	10.4	40.9	62.4	37.4	36.7	47.3	36.5	38.3	29.0	41.3	30.6	63.5	64.2	54.6	36.0	40.0	40.5	41.8	40.2	22.3
3	12.5	38.2	48.3	37.2	41.1	49.1	37.8	39.8	31.0	42.8	27.5	54.7	58.3	52.5	34.2	48.2	39.7	40.1	37.5	24.5
MAGIC x M4																				
4	1.2	36.4	56.4	19.1	24.6	37.7	39.8	22.9	14.0	25.4	19.4	24.4	32.2	19.9	34.7	21.6	19.8	53.6	19.9	30.7

4.3.5 Candidate Genes Involved in Seed Abortion Response

A total of 624 genes were present under the four QTL with 93 genes overlapping between two of the QTL for MAGIC x C4 which were in close proximity on chromosome 3 (Table 4.7). Hence, 531 unique genes were obtained of which 245 protein coding genes of known function were retained for further screening.

Any gene that had functions related to pollen growth and maturation, cytoskeletal organisation, cell division, cell differentiation and seed development were shortlisted. In addition to these, auxin related genes were also shortlisted as impaired auxin transport in the endosperm has been shown to be involved in seed abortion (Figueiredo *et al.*, 2016).

Key genes related to auxin biosynthesis, binding, transport, or response found under the QTLs included *CUL1* (AT4G02570), *ABP1* (AT4G02980) and *AFB1* (AT4G03190) on chromosome 4, and *PAD3* (AT3G26830), *SLK1* (AT3G26900), *PMZ* (AT3G28210) and *ABCB15*, *ABCB16* and *ABCB17* (AT3G28345, AT3G28360 and AT3G28380 respectively) on chromosome 3. Other genes under the QTLs known to regulate seed development included *SEC15B* (AT4G02350), *ATFH* (AT4G03240), *RST1* (AT3G27670), *GL1* (AT3G27920). A total of 11 and 16 candidate genes that could potentially be involved in regulating seed abortion in MAGIC x M4 crosses and MAGIC x C4 crosses respectively were shortlisted. The candidate genes are listed in Table 4.8 and Table 4.9 for M4 QTL and C4 QTL respectively.

Table 4.7 | Tally of the different types of genes present in a ± 250kb around each QTL marker.

Type of Genes	Total Number
Total genes	624
Overlapping genes	93
Other RNA genes	3
Pre-tRNA genes	5
Pseudogenes	23
Small nucleolar RNA genes (snoRNA)	3
Transposable element genes	108
Genes with unknown functions	76
Genes coding unknown proteins	64
Protein-coding Gene with known functions	245

Table 4.8 | Candidate Genes under the MAGIC x M4 seed abortion QTL.

Gene	Auxin Related	Pollen Development	Cytoskeleton	Mitosis / Cell cycle / Cell Division	Seed Development	Differentiation
AT4G02350		pollen tube growth; pollen germination				
AT4G02460		pollen development	sister chromatid cohesion	mitotic cell cycle; cell division; regulation of DNA replication	embryonic pattern specification; seed dormancy process; seed development; embryo development ending in seed dormancy; seed maturation	embryo sac egg cell differentiation; regulation of cell differentiation
AT4G02570	response to auxin			mitotic cell cycle; cell division	seed development; seed maturation; embryonic pattern specification; embryo development ending in seed dormancy; seed dormancy process	regulation of cell differentiation
AT4G02590					double fertilization forming a zygote and endosperm	
AT4G02640					positive regulation of seed maturation	
AT4G02820						embryo sac egg cell differentiation
AT4G02980	auxin binding			positive regulation of cell division		
AT4G03190	auxin binding; response to auxin					
AT4G03190		pollen maturation				
AT4G03240				cell division		
AT4G03270				mitotic cell cycle		

Table 4.9 | Candidate Genes under the MAGIC x C4 seed abortion QTL.

Gene	Auxin Related	Pollen Development	Cytoskeleton	Mitosis / Cell cycle / Cell Division	Seed Development	Differentiation
AT3G26830	auxin biosynthesis					
AT3G26870		pollen tube growth				
AT3G26900	auxin biosynthesis					cell differentiation
AT3G27000			actin cytoskeleton organization			
AT3G27060				DNA-dependent DNA replication		
AT3G27530					post-embryonic development	
AT3G27640				DNA-dependent DNA replication		
AT3G27660					embryo development ending in seed dormancy	
AT3G27670			sister chromatid cohesion	cell division	seed dormancy process; seed maturation; embryo development ending in seed dormancy; embryonic pattern specification	regulation of cell differentiation
AT3G27730			sister chromatid cohesion	mitotic nuclear division		
AT3G27920						epidermal cell differentiation
AT3G27960		pollen tube growth				
AT3G28210	response to auxin					
AT3G28345	auxin efflux					
AT3G28360	auxin efflux					
AT3G28380	auxin efflux				seed maturation	

4.4 DISCUSSION

4.4.1 Natural variation was seen in Arabidopsis thaliana interploidy crosses

Polyploidy is often considered a speciation mechanism as it can cause reproductive isolation from diploids due to the presence of post-zygotic hybridization barriers in interploidy crosses. Here, we use *A. thaliana* as a model to understand these barriers by investigating the role of maternal genotypes, paternal genotypes, and maternal-by-paternal genotype interaction on the proportion of triploid F1 seeds aborted resulting from interploidy crosses between diploid mothers and tetraploid fathers. Extensive research done in the past on parent-of-origin specific effects have greatly advanced the mechanistic understanding of the molecular basis of seed abortion and seed development in general (Dilkes & Comai, 2004; Schatlowski & Kohler, 2012; Kradolfer *et al.*, 2013), however, the imprinted genes involved in seed abortion are likely to be a subset of all genes that could be responsible for interploidy seed failure. Hence, continuing to look at parental genotypes and their effects will complement understanding of the genetic basis of the interploidy hybridisation effects on seed abortion.

Based on our findings from the pilot experiment we observed that there was greater seed abortion when each of the seven maternal genotypes were crossed to Col pollen in comparison to Ler pollen (Figure 4.2). Our observations are in line with previous work by Dilkes *et al.*, (2008) who presented one of the earliest studies of paternal genotype-dependent seed abortion response. They found that paternal excess interploidy crosses with *A. thaliana* genotype Col as a tetraploid pollen donor resulted in a strong seed abortion response that was not seen with tetraploid pollen from Ler and C24 genotypes of *A.thaliana*, suggesting that the response to interploidy hybridization was dependent on the paternal genotype. In addition to this, we also found significant maternal variation across the genotypes (Table 4.1), suggesting that the maternal genotype also contribute towards the interploidy cross response. We observed that individually, both maternal and paternal genotypes had significant effects on triploid seed abortion. Surprisingly, there haven't been many studies that have looked into understanding maternal variation and maternal-by-paternal genotype interaction, and thus, our understanding of the role that parental genotypes *per se* play in interploidy hybridisation remain limited. We also observed that the magnitude of

difference in percentage seed abortion between crosses with the two paternal tetraploid genotypes varied from 7.7% in Wil mothers to 75.1% in Col mothers (Figure 4.2) suggesting that seed abortion might not be passively regulated by maternal or paternal genotypes independently but may work in combination to determine the final seed abortion response. Due to the low sample size for the pilot experiment, we were unable to determine if the maternal-by-paternal interaction in determining seed abortion response was significant, however, there are indications that this may be the case. An experiment with a larger sample size and multiple replicates of cross-pairs between several genotypes was conducted to detect any interaction between parental genotypes.

A similar pattern was seen in the larger scale (main experiment) using the MAGIC lines (Table 4.4), where we observed significant maternal and paternal effects on percentage seed abortion of the F1 hybrid, but did not find any significant maternal-by-paternal genotype interaction. In this case, however, the failure to see a significant interaction may be due to the limited number of paternal genotypes studied (Col and Mt). If both paternal genotypes have similar genetic mechanisms underlying the seed abortion response, it is likely that they would interact with the maternal genotypes in a similar manner, thus resulting in nonsignificant interaction effects. Although we do not find significant interaction effects in the pilot or main experiments, this should not be interpreted as evidence of absence of interaction effects. We recommend expanding the combinations and replicates of maternal and paternal genotype crosses to provide more robust data before interpreting and ruling out maternal-by-paternal genotype interaction effects.

4.4.2 Independent genetic mechanisms exist depending on the paternal genotype of the interploidy cross

The paternal genotype effects on seed abortion that we saw statistically are also backed up genetically by the detection of non-overlapping QTL peaks for MAGIC x C4 and MAGIC x M4 that crossed the significance threshold. We found three significant QTL peaks for the MAGIC x C4 percentage seed abortion phenotype that were in close proximity on chromosome 3. Only one QTL was detected for MAGIC x M4 percentage seed abortion phenotype on chromosome 4.

On chromosome 2, there appear to be a peak above suggestive threshold (albeit not significant) that appears to be in the same region that has been previously identified as a major locus affecting seed survival known as *DSL1* (Dilkes *et al.*, 2008). It is likely that in future experiments with increased replication and using a wider set of MAGIC lines will reduce non-genetic variance and improve the statistical power, enhancing the ability to detect more QTL. Our findings suggest that further exploration of seed abortion responses using MAGIC lines can be a powerful tool to provide invaluable information and uncover new loci on the genome that respond to paternal excess tetraploid crosses in a genotype-specific manner.

4.4.3 The candidate genes shortlisted may have a role in triploid block responses

We scanned for candidate genes in close proximity to the QTL markers. A number of genes with a range of functions that could potentially be involved in maternal genotype response to C4 and M4 crosses were shortlisted (Table 4.8-4.9). These genes have known functions directly or indirectly related to seed development.

Several genes from those enlisted have been shown to directly regulate seed abortion e.g *ABP1* mutants exhibit cell division and elongation defects resulting in embryo lethality (Chen *et al.*, 2001) and *SLK1* functions redundantly with *SEU* and *SLK2* in early embryo development (Bao *et al.*, 2010). *SEC15B*, that encodes an exocyst complex component subunit required in the stigma to facilitate pollen hydration (Hála *et al.*, 2008; Chapman & Goring, 2010); *ATFH* knockouts exhibit an embryo-lethal phenotype (Vazzola *et al.*, 2007) thought to be due to inadequate utilization of iron reserves essential for embryogenesis (Jain *et al.*, 2019); *RST1* regulates lipid synthesis of cuticular wax and *RST1* mutants have been shown to produce shrunken non-viable seeds (Chen *et al.*, 2005); AT4G02820 encodes a pentatricopeptide repeat-containing protein (PPR), a family of proteins that play a non-redundant role in embryogenesis by regulating differentiation of the transfer cells in the basal endosperm (Gutiérrez-Marcos *et al.*, 2007; Chen *et al.*, 2018); *GL1*, primarily known to play a central role in trichome development, is also known to interact with *TTG1* (Larkin *et al.*, 1999) that is responsible for regulating epidermal patterning during embryogenesis (Lin & Schiefelbein, 2001) and for the regulation of *TTG2* – a maternally expressed transcription factor shown to be

involved in F1 lethality in interploidy crosses (Garcia *et al.*, 2005; Dilkes *et al.*, 2008; Gonzalez *et al.*, 2016).

Maternal variation for one or more of these genes could be linked to a host of post zygotic changes consequently leading to maternal variation in seed abortion. Further investigation of the genes will be needed in order to validate their role in seed abortion. A single gene or several of the enlisted genes could work either in isolation or more likely as part of a larger pathway to determine seed abortion responses in interploidy crosses, the understanding of which can advance our knowledge of the genetic mechanisms that underlie post zygotic hybridization barriers. As most quantitative traits are polygenic, a single gene in unlikely to explain the causal effects for a complex phenotype, like the triploid block. Hence, exploring the roles of several contributing factors affecting a phenotype will enhance our understanding of complex networks that determining traits like seed abortion.

4.4.4 MAGIC lines as a means to understand genetic mechanisms involved in interploidy hybridization

This represents the first study wherein the MAGIC lines have been used to study interploidy hybridization barriers by identifying QTL locations controlling seed abortion phenotype in paternal excess interploidy crosses. Two-parent populations have been used in the past and have revealed one large effect QTL (Dilkes *et al.*, 2008). In comparison to RILs or F2 backcross (BC), multi-parent populations are an advanced tool to provide greater allelic diversity and better QTL resolution due to increased recombination in small chromosomal regions that can be used for fine-mapping (Kover *et al.*, 2009; Bandillo *et al.*, 2013).

The *A. thaliana* MAGIC lines can be an efficient and elegant way of understanding interploidy hybridization by providing larger genetic and phenotypic diversity than previously studied. As the lines do not require repeated genotyping, a larger number of replicates for various cross-combinations can be crossed and data can be accumulated over an extended period of time, providing practical feasibility (Kover *et al.*, 2009). By using a wider range of MAGIC lines as well as paternal tetraploid genotypes, this approach has the potential to enhance elucidation of the genetic architecture underlying a range of triploid block mechanisms corresponding to the natural genetic variation seen for the trait.

The understanding of pathways involved in inter-ploidy hybridization barriers can also provide hints to pathways that might be involved in inter-species hybridization barriers that can have implications on our understanding evolutionary processes of speciation and commercial applications in the improvement of agronomically important traits.

Chapter 5 : **CONCLUSIONS**

5.1 Key findings from the three data chapters

5.1.1 The responses to increased ploidy are genotype-specific

In chapter 2, we observed that significant genotype-by-ploidy effects were seen for all phenotypic traits studied. A closer inspection of these effects revealed that there are likely to be two mechanisms that interact to determine the ultimate phenotype (i) Nucleotypic effects that are genotype-independent that result in common responses to increased ploidy in all genotypes; (ii) Genotype-dependent effects that result in responses to ploidy based on the specific genotype under study.

For all traits, significant genotype-by-ploidy interaction effects were observed. Depending on the phenotypic trait being studied, this can have several interpretations (i) For traits in which common responses to increased ploidy were observed for all genotypes, it implied that the scale (or extent) of the response was genotype-dependent (nucleotype + quantitative genotype effects); (ii) For traits in which only some genotypes showed responses to increased ploidy, it implied that the genotype determined if there would be any response to ploidy or not (qualitative genotype effects) and what the scale of the response would be (quantitative genotype effects). It can be seen, that in either case, genotypes play a central role in determining the overall response.

We also observed that no single genotype responded to all ploidy-related phenotypes. Depending on the trait, different genotypes exhibited significant responses to increased ploidy. This suggests that there is no single pathway that if activated in a particular genotype would lead to polyploidy responses for all traits, or in other words, genotypes *per se* are not susceptible (responsive) or tolerant (unresponsive) to ploidy increases, their responses to increased ploidy depend on the phenotype being looked at.

In chapter 3, we observed that gene expression changes between ploidy levels were genotype-specific. A larger number of genes were differentially expressed in response to ploidy in Sf relative to Ler. This intuitively suggests that ploidy induced changes in Ler are moderate and Sf shows a stronger response. However, when looking at the phenotypes for the two genotypes, both genotypes showed significant changes. This could mean that (i) Ler had few genes of large-effects that were differentially expressed, thus changes in only a few genes could lead to significant phenotypic

alterations; and in comparison Sf had many genes of small-effects that changed expression, cumulatively altering phenotypes; (ii) there was more variation in gene expression among Ler replicates and hence most genes didn't pass the statistical analysis thresholds. Based on the current data, it is difficult to make a decision as to which case is more likely. However, given that studies in the past have also reported a small number of differentially expressed genes in Ler, coupled with our observation that Ler samples of higher ploidy were strongly enriched for circadian clock genes and pathways (which is considered to be a master-regulator of several processes), the former scenario seems more likely.

Polyploidy changed the expression of genes belonging to different pathways for the two genotypes. Ler was enriched for circadian clock pathway whereas Sf was enriched for metabolic pathways. This suggested that despite exhibiting relatively similar phenotypic effects for most traits, the underlying mechanisms that controlled them were very different and was dependant on the genotype.

In chapter 4, we found that a large range of percentage seed abortion spanning from 0-100% was observed in inter-ploidy crosses between the MAGIC lines and the two paternal tetraploid genotypes. It was seen that there was low correlation in percentage seed abortion for the same set of MAGIC lines crossed to Col tetraploids and Mt tetraploids. This was supported by the fact that QTL for percentage seed abortion with Mt tetraploid pollen was on a different chromosome to the QTL detected for crosses with Col tetraploid pollen. This suggested that although crosses with Mt tetraploids and Col tetraploids both resulted in similar average percentage seed abortions over all the crosses, their genetic mechanisms controlling the response were different and dependent on the genotype. We also found that maternal and paternal genotypes both contributed to the seed abortion response.

5.1.2 Polyploidy series are a good way to see ploidy-related patterns

As with experiments that use time-series or concentration gradients to explain how time or concentration affects a trait in question, having multiple reference points for ploidy helps identify patterns that may otherwise go unnoticed. It also allows us to understand if the responses observed scale linearly with increasing ploidy or if there are thresholds that restrict them.

In chapter 3, Using diploids, tetraploids and octoploids, we found that the phenotypic responses between tetraploids and octoploids followed the same direction as the responses between diploids and tetraploids. The magnitude was not always proportional, suggesting that phenotypic responses do not scale linearly as ploidy increases.

In transcriptome studies in particular, we found that the ability to detect differentially expressed genes was enhanced by adding a third ploidy level. Given that it is generally difficult to distinguish true transcriptomic changes from background noise, resulting in the detection of a small number of differentially expressed genes, this approach provides a good solution as the increase in difference between the ploidy levels (a difference of 4 units between tetraploids and octoploids and a difference of 6 units between diploids and octoploids, as opposed to 2 between diploids and tetraploids) appears to enhance the response to ploidy increase, making differences in gene expression patterns more evident.

5.1.3 Colchicine has heritable effects that need to be controlled for

In chapters 2 and 3, we observed that colchicine had heritable effects on phenotype and gene expression respectively. We found that when generating polyploid lines via colchicine, a large proportion of the treated lines do not undergo genome doubling. As these lines have been exposed to colchicine, they serve as more appropriate controls to compare colchicine treated neopolyploid lines with.

We further find that colchicine effects are reproducible within independently generated lines, further reiterating the need to use colchicine treated lines as controls as the changes they may bring about, albeit subtle, appear to be non-stochastic.

5.2 Future directions

Our data provides evidence of significant genotype effects in response to increased ploidy for phenotypic traits, transcriptomic changes as well as interploidy seed abortion responses. Given that the nature and scale of the response to increased ploidy is largely determined by the genetic background of the plant, it highlights the need for future studies to include the use of multiple genotypes in experimental designs when studying ploidy effects. Using an expanding set of genotypes can help reveal multiple underlying pathways that regulate ploidy responses, which will provide a more

comprehensive overview of processes in neoautopolyploids. Our current understanding of polyploidy processes for any particular species is based on a limited number of genotypes, which perhaps does not represent the complete picture. By accruing data across multiple genotypes in multiple species, inferences can be pooled and key genes, pathways and networks regulating ploidy responses can be identified.

The effect that the method of inducing polyploidy may have on our interpretation of ploidy responses are oftentimes neglected. At present, most studies compare synthetic neoautopolyploids to untreated diploids, findings from which are often confounded effects of ploidy effects and colchicine effects. Thus, to isolate ploidy effects, future studies should use colchicine exposed diploids as controls.

Finally, incorporating multiple ploidy levels in future experiments can be helpful as they aid the understanding of increased ploidy responses by providing an additional point of reference, which makes elucidation of phenotypic or transcriptomic patterns prominent.

Chapter 6 **: REFERENCES**

Adams S, Vinkenoog R, Spielman M, Dickinson HG, Scott RJ. **2000**. Parent-of-origin effects on seed development in *Arabidopsis thaliana* require DNA methylation. *Development* **127**: 2493–2502.

Ainouche ML, Baumel A, Salmon A, Yannic G. **2004**. Hybridization, polyploidy and speciation in *Spartina* (Poaceae). *New Phytologist* **161**: 165–172.

Albertin W, Brabant P, Catrice O, Eber F, Jenczewski E, Chèvre A-M, Thiellement H. **2005**. Autopolyploidy in cabbage (*Brassica oleracea* L.) does not alter significantly the proteomes of green tissues. *Proteomics* **5**: 2131–2139.

Albertin W, Marullo P. **2012**. Polyploidy in fungi: Evolution after whole-genome duplication. *Proceedings of the Royal Society B: Biological Sciences* **279**: 2497–2509.

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JS (Pat). **2017**. Polyploidy and interspecific hybridization: Partners for adaptation, speciation and evolution in plants. *Annals of Botany* **120**: 183–194.

Allario T, Brumos J, Colmenero-Flores JM, Iglesias DJ, Pina JA, Navarro L, Talon M, Ollitrault P, Morillon R. **2013**. Tetraploid Rangpur lime rootstock increases drought tolerance via enhanced constitutive root abscisic acid production. *Plant, Cell and Environment* **36**: 856–868.

Allario T, Brumos J, Colmenero-Flores JM, Tadeo F, Froelicher Y, Talon M, Navarro L, Ollitrault P, Morillon R. **2011**. Large changes in anatomy and physiology between diploid Rangpur lime (*Citrus limonia*) and its autotetraploid are not associated with large changes in leaf gene expression. *Journal of Experimental Botany* **62**: 2507–2519.

Allum JF, Bringloe DH, Roberts A V. **2007**. Chromosome doubling in a *Rosa rugosa* Thunb. hybrid by exposure of in vitro nodes to oryzalin: the effects of node length, oryzalin concentration and exposure time. *Plant Cell Reports* **26**: 1977–1984.

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M. **2009**. What has natural variation taught us about plant development, physiology, and adaptation? *The Plant Cell* **21**: 1877–1896.

Alonso-Blanco C, Koornneef M. **2000**. Naturally occurring variation in *Arabidopsis*: An underexploited resource for plant genetics. *Trends in Plant Science* **5**: 22–29.

Amah D, van Biljon A, Maziya-Dixon B, Labuschagne M, Swennen R. **2019**. Effects of in vitro polyploidization on agronomic characteristics and fruit carotenoid content; Implications for banana genetic improvement. *Frontiers in Plant Science* **10**: 1450.

Ansari N, Thomas H. **1983**. A study of homoeologous relationships in the cultivated oat *Avena sativa* (2n=6x=42). *Theoretical and Applied Genetics* **66**: 303–305.

Anssour S, Krügel T, Sharbel TF, Saluz HP, Bonaventure G, Baldwin IT. **2009**. Phenotypic,

genetic and genomic consequences of natural and synthetic polyploidization of *Nicotiana attenuata* and *Nicotiana obtusifolia*. *Annals of botany* **103**: 1207–1217.

Ari E, Djapo H, Mutlu N, Gurbuz E, Karaguzel O. **2015**. Creation of variation through gamma irradiation and polyploidization in *Vitex agnus-castus* L. *Scientia Horticulturae* **195**: 74–81.

Arrigo N, Barker MS. **2012**. Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology* **15**: 140–146.

Ascough GD, Van Staden J, Erwin JE. **2008**. Effectiveness of colchicine and oryzalin at inducing polyploidy in *Watsonia lepida* N.E. Brown. *Horticultural Science* **43**: 2248–2251.

Aversano R, Caruso I, Aronne G, De Micco V, Scognamiglio N, Carputo D. **2013**. Stochastic changes affect *Solanum* wild species following autopolyploidization. *Journal of experimental botany* **64**: 625–635.

Aversano R, Scarano M-T, Aronne G, Caruso I, D'Amelia V, De Micco V, Fasano C, Termolino P, Carputo D. **2015**. Genotype-specific changes associated to early synthesis of autotetraploids in wild potato species. *Euphytica* **202**: 307–316.

Bandillo N, Raghavan C, Muyco PA, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, Tung C-W, McCouch S, Thomson M, Mauleon R, *et al.* 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. *Rice* **6**: 11.

Bao F, Azhakanandam S, Franks RG. **2010**. *SEUSS* and *SEUSS-LIKE* transcriptional adaptors regulate floral and embryonic development in Arabidopsis. *Plant physiology* **152**: 821–836.

Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. **2016**. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist* **210**: 391–398.

Barkla BJ, Rhodes T, Tran K-NT, Wijesinghege C, Larkin JC, Dassanayake M. **2018**. Making epidermal bladder cells bigger: Developmental- and salinity-induced endopolyploidy in a model halophyte. *Plant physiology* **177**: 615–632.

Barow M. **2006**. Endopolyploidy in seed plants. *BioEssays* **28**: 271–281.

Bendix C, Marshall CM, Harmon FG. **2015**. Circadian clock genes universally control key agricultural traits. *Molecular Plant* **8**: 1135–1152.

Bennett MD. **1971**. The duration of meiosis. *Proceedings of the Royal Society B: Biological Sciences* **178**: 277–299.

Bennett MD. **2004**. Perspectives on polyploidy in plants – ancient and neo. *Biological Journal of the Linnean Society* **82**: 411–423.

Bennett MD, Riley R. **1972**. Nuclear DNA content and minimum generation time in herbaceous plants. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **181**: 109–135.

Bergström I. **1938**. Tetraploid apple seedlings obtained from the progeny of triploid varieties. *Hereditas* **24**: 210–215.

Bergström I. **1940**. On the progeny of diploid x triploid *Populus tremula*. *Hereditas* **26**: 191–201.

Bertrand B, Bardil A, Baraille H, Dussert S, Doulbeau S, Dubois E, Severac D, Dereeper A, Etienne H. **2015**. The greater phenotypic homeostasis of the allopolyploid *Coffea arabica* improved the transcriptional homeostasis over that of both diploid parents. *Plant & cell physiology* **56**: 2035–2051.

Bevan M, Walsh S. **2005**. The *Arabidopsis* genome : A foundation for plant research. *Genome Research* **15**: 1632–1642.

Bhosale R, Boudolf V, Cuevas F, Lu R, Eekhout T, Hu Z, Van Isterdael G, Lambert GM, Xu F, Nowack MK, *et al.* 2018. A spatiotemporal DNA endoploidy map of the *Arabidopsis* root reveals roles for the endocycle in root development and stress adaptation. *The Plant Cell* **30**: 2330–2351.

Blanc G, Wolfe KH. **2004a**. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *The Plant Cell* **16**: 1679–1691.

Blanc G, Wolfe KH. **2004b**. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell* **16**: 1667–1678.

Bolger AM, Lohse M, Usadel B. **2014**. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bomblies K, Madlung A. **2014**. Polyploidy in the *Arabidopsis* genus. *Chromosome Research* **22**: 117–134.

Borevitz JO, Nordborg M. **2003**. The impact of genomics on the study of natural variation in *Arabidopsis*. *Plant Physiology* **132**: 718–725.

Bramsiepe J, Wester K, Weinl C, Roodbarkelari F, Kasili R, Larkin JC, Hülskamp M, Schnittger A. **2010**. Endoreplication controls cell fate maintenance. *PLoS genetics* **6**: e1000996–e1000996.

Bray NL, Pimentel H, Melsted P, Pachter L. **2016**. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* **34**: 525–527.

Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA. **2019**. R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* **211**: 495–502.

Burton TL, Husband BC. **2001**. Fecundity and offspring ploidy in matings among diploid, triploid and tetraploid *Chamerion angustifolium* (Onagraceae): Consequences for tetraploid establishment. *Heredity* **87**: 573–582.

Bushell C, Spielman M, Scott RJ. **2003**. The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. *The Plant Cell* **15**: 1430–1442.

Camargo A V, Mackay I, Mott R, Han J, Doonan JH, Askew K, Corke F, Williams K, Bentley AR. **2018**. Functional mapping of Quantitative Trait Loci (QTLs) associated with plant performance in a wheat MAGIC mapping population. *Frontiers in plant science* **9**: 887.

Caperta AD, Delgado M, Ressurreição F, Meister A, Jones RN, Viegas W, Houben A. **2006**. Colchicine-induced polyploidization depends on tubulin polymerization in c-metaphase cells. *Protoplasma* **227**: 147–153.

Čertner M, Sudová R, Weiser M, Suda J, Kolář F. **2019**. Ploidy-altered phenotype interacts with local environment and may enhance polyploid establishment in *Knautia serpentinicola* (Caprifoliaceae). *New Phytologist* **221**: 1117–1127.

Chapman LA, Goring DR. **2010**. Pollen–pistil interactions regulating successful fertilization in the Brassicaceae. *Journal of Experimental Botany* **61**: 1987–1999.

Chen ZJ. **2007**. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* **58**: 377–406.

Chen C, E Z, Lin H-X. **2016**. Evolution and molecular control of hybrid incompatibility in plants. *Frontiers in Plant Science* **7**: 1208.

Chen X, Goodwin SM, Liu X, Chen X, Bressan RA, Jenks MA. **2005**. Mutation of the *RESURRECTION1* locus of *Arabidopsis* reveals an association of cuticular wax with embryo development. *Plant Physiology* **139**: 909–919.

Chen ZJ, Ni Z. **2006**. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* **28**: 240–252.

Chen JG, Ullah H, Young JC, Sussman MR, Jones AM. **2001**. *ABP1* is required for organized cell elongation and division in *Arabidopsis* embryogenesis. *Genes & development* **15**: 902–911.

Chen G, Zou Y, Hu J, Ding Y. **2018**. Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. *BMC genomics* **19**: 720.

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev E V, Leitch AR, Soltis PS, Soltis DE. **2012**. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proceedings of the National Academy of Sciences* **109**: 1176–1181.

Clarindo WR, Carvalho CR. **2008**. First *Coffea arabica* karyogram showing that this species is a true allotetraploid. *Plant Systematics and Evolution* **274**: 237.

Cochran JJ. **2019**. What is the bootstrap? *Significance* **16**: 8–9.

Comai L. **2005**. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**: 836–846.

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. **2000**. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *The Plant cell* **12**: 1551–1568.

Conant GC, Wolfe KH. **2008**. Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics* **9**: 938–950.

Coolen S, Van Pelt JA, Van Wees SCM, Pieterse CMJ. **2019**. Mining the natural genetic variation in *Arabidopsis thaliana* for adaptation to sequential abiotic and biotic stresses. *Planta* **249**: 1087–1105.

Corneillie S, De Storme N, Van Acker R, Fangel JU, De Bruyne M, De Rycke R, Geelen D, Willats WGT, Vanholme B, Boerjan W. **2019**. Polyploidy affects plant growth and alters cell wall composition. *Plant Physiology* **179**: 74–87.

Costa LM, Gutièrrez-Marcos JF, Dickinson HG. **2004**. More than a yolk: The short life and complex times of the plant endosperm. *Trends in Plant Science* **9**: 507–514.

Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL. **2008**. Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome biology* **9**: R130–R130.

Cuñado N, Blazquez S, Melchor L, Pradillo M, Santos JL. **2005**. Understanding the cytological diploidization mechanism of polyploid wild wheats. *Cytogenetic and Genome Research* **109**: 205–209.

D'Amelia V, Aversano R, Ruggiero A, Batelli G, Appelhagen I, Dinacci C, Hill L, Martin C, Carputo D. **2018**. Subfunctionalization of duplicate *MYB* genes in *Solanum commersonii* generated the cold-induced *ScAN2* and the anthocyanin regulator *ScAN1*. *Plant, Cell & Environment* **41**: 1038–1051.

Datta AK, Mandal A, Das D, Gupta S, Saha A, Paul R, Sengupta S, Halder S, Biswas S. **2016**. Polyploidy in angiosperms: Genetic insight to the phenomenon. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences* **86**: 513–522.

Davidson D. **1961**. Mechanisms of reorganization and cell repopulation in meristems in roots of *Vicia faba* following irradiation and colchicine. *Chromosoma* **12**: 484–504.

Del-Pozo JC, Ramirez-Parra E. **2014**. Deciphering the molecular bases for drought tolerance in *Arabidopsis* autotetraploids. *Plant, Cell and Environmentl* **37**: 2722–2737.

Dempewolf H, Baute G, Anderson J, Kilian B, Smith C, Guarino L. **2017**. Past and future use of wild relatives in crop breeding. *Crop Science* **57**: 1070–1082.

Denaeghel HERR, Van Laere K, Leus L, Lootens P, Van Huylenbroeck J, Van Labeke M-CC. **2018**. The variable effect of polyploidization on the phenotype in *Escallonia*. *Frontiers in Plant Science* **9**: 1–17.

Dhooghe E, Van Laere K, Eeckhaut T, Leus L, Van Huylenbroeck J. **2011**. Mitotic chromosome doubling of plant tissues in vitro. *Plant Cell, Tissue and Organ Culture (PCTOC)* **104**: 359–373.

Dilkes BP, Comai L. **2004**. A differential dosage hypothesis for parental effects in seed development. *The Plant Cell* **16**: 3174–3180.

Dilkes BP, Spielman M, Weizbauer R, Watson B, Burkart-Waco D, Scott RJ, Comai L.

2008. The maternally expressed WRKY transcription factor *TTG2* controls lethality in interploidy crosses of *Arabidopsis*. *PLoS Biology* **6**: e308.

Ding M, Chen ZJ. **2018**. Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. *Current opinion in plant biology* **42**: 37–48.

Dodsworth S, Chase MW, Leitch AR. **2016**. Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Botanical Journal of the Linnean Society* **180**: 1–5.

Doebley JF, Gaut BS, Smith BD. **2006**. The molecular genetics of crop domestication. *Cell* **127**: 1309–1321.

Doležel J, Binarová P, Lcretti S. **1989**. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum* **31**: 113–120.

Doyle JJ, Coate JE. **2019**. Polyploidy, the nucleotype, and novelty: The impact of genome goubling on the biology of the cell. *International Journal of Plant Sciences* **180**: 1–52.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. **2008**. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443–461.

Doyle JJ, Sherman-Broyles S. **2017**. Double trouble: Taxonomy and definitions of polyploidy. *New Phytologist* **213**: 487–493.

Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires PK, Leebens-Mack J, dePamphilis CW. **2010**. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* **10**: 61.

Dun X, Shen W, Hu K, Zhou Z, Xia S, Wen J, Yi B, Shen J, Ma C, Tu J, *et al.* 2014. Neofunctionalization of duplicated *Tic40* genes caused a gain-of-function variation related to male fertility in *Brassica oleracea* lineages. *Plant physiology* **166**: 1403–1419.

Duruflé H, Ranocha P, Mbadinga Mbadinga DL, Déjean S, Bonhomme M, San Clemente H, Viudes S, Eljebbawi A, Delorme-Hinoux V, Sáez-Vásquez J, *et al.* 2019. Phenotypic trait variation as a response to altitude-related constraints in *Arabidopsis* populations. *Frontiers in Plant Science* **10**: 430.

Duszynska D, Vilhjalmsson B, Castillo Bravo R, Swamidatta S, Juenger TE, Donoghue MTA, Comte A, Nordborg M, Sharbel TF, Brychkova G, *et al.* 2019. Transgenerational effects of inter-ploidy cross direction on reproduction and F2 seed development of *Arabidopsis thaliana* F1 hybrid triploids. *Plant Reproduction* **32**: 275–289.

Edgar BA, Orr-Weaver TL. **2001**. Endoreplication cell cycles: More for less. *Cell* **105**: 297–306.

Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, Smith RD, Teresi SJ, Nelson ADL, Wai CM, *et al.* 2019. Origin and evolution of the octoploid strawberry genome. *Nature Genetics* **51**: 541–547.

Efron B. **1979**. Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**: 1–26.

Efron B, Hastie T. **2016**. *Computer Age Statistical Inference*. Cambridge University Press.

Eliášová A, Münzbergová Z. **2014**. Higher seed size and germination rate may favour autotetraploids of *Vicia cracca* L. (Fabaceae). *Biological Journal of the Linnean Society* **113**: 57–73.

Erilova A, Brownfield L, Exner V, Rosa M, Twell D, Scheid OM, Hennig L, Köhler C. **2009**. Imprinting of the Polycomb group gene *MEDEA* serves as a ploidy sensor in *Arabidopsis*. *PLoS Genetics* **5**.

Fasano C, Diretto G, Aversano R, D'Agostino N, Di Matteo A, Frusciante L, Giuliano G, Carputo D. **2016**. Transcriptome and metabolome of synthetic *Solanum* autotetraploids reveal key genomic stress events following polyploidization. *New Phytologist* **210**: 1382–1394.

Figueiredo DD, Batista RA, Roszak PJ, Hennig L, Köhler C. **2016**. Auxin production in the endosperm drives seed coat development in *Arabidopsis*. *eLife* **5**: 1–23.

Figueiredo DD, Batista RA, Roszak PJ, Kohler C. **2015**. Auxin production couples endosperm development to fertilization. *Nature Plants* **1**: 15184.

Florez-Rueda AM, Paris M, Schmidt A, Widmer A, Grossniklaus U, Städler T. **2016**. Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Molecular Biology and Evolution* **33**: 2935–2946.

Francis A, Jones RN. **1989**. Heritable nature of colchicine induced variation in diploid *Lolium perenne*. *Heredity* **62**: 407–410.

Francis A, Jones RN, Parker JS, Posselt UK. **1990**. Colchicine-induced heritable variation in cell size and chloroplast numbers in leaf mesophyll cells of diploid ryegrass (*Lolium perenne* L.). *Euphytica* **49**: 49–55.

Freeling M. **2009**. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology* **60**: 433–453.

Freeling M, Scanlon MJ, Fowler JE. **2015**. Fractionation and subfunctionalization following genome duplications: Mechanisms that drive gene content and their consequences. *Current Opinion in Genetics & Development* **35**: 110–118.

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. **2007**. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *The Plant cell* **19**: 3403–3417.

Galbraith DW, Harkins KR, Knapp S. **1991**. Systemic endopolyploidy in *Arabidopsis thaliana*. *Plant Physiology* **96**: 985–989.

Gallagher JP, Grover CE, Hu G, Wendel JF. **2016**. Insights into the ecology and evolution of polyploid plants through network analysis. *Molecular Ecology* **25**: 2644–2660.

Gallardo MH, González CA, Cebrián I. **2006**. Molecular cytogenetics and allotetraploidy in the red vizcacha rat, *Tympanoctomys barrerae* (Rodentia, Octodontidae). *Genomics* **88**: 214–221.

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, *et al.* 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* **477**: 419–423.

Gao R, Wang H, Dong B, Yang X, Chen S, Jiang J, Zhang Z, Liu C, Zhao N, Chen F. **2016**. Morphological, genome and gene expression changes in newly induced autopolyploid *Chrysanthemum lavandulifolium* (Fisch. ex Trautv.) Makino. *International Journal of Molecular Sciences* **17**: 1690.

Garcia D, Gerald JNF, Berger F. **2005**. Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis*. *The Plant Cell* **17**: 52–60.

Garner AG, Kenney AM, Fishman L, Sweigart AL. **2016**. Genetic loci with parent-of-origin effects cause hybrid seed lethality in crosses between *Mimulus* species. *New Phytologist* **211**: 319–331.

Gegas VC, Wargent JJ, Pesquet E, Granqvist E, Paul ND, Doonan JH. **2014**. Endopolyploidy as a potential alternative adaptive strategy for *Arabidopsis* leaf size variation in response to UV-B. *Journal of experimental botany* **65**: 2757–2766.

Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK. **2013**. CHH islands: De novo DNA methylation in near-gene chromatin regulation in maize. *Genome research* **23**: 628–637.

Godfree RC, Marshall DJ, Young AG, Miller CH, Mathews S. **2017**. Empirical evidence of fixed and homeostatic patterns of polyploid advantage in a keystone grass exposed to drought and heat stress. *Royal Society open science* **4**: 170934.

Goldblatt P. **1980**. Uneven diploid chromosome numbers and complex heterozygosity in *Homeria* (Iridaceae). *Systematic Botany* **5**: 337–340.

Gonzalez A, Brown M, Hatlestad G, Akhavan N, Smith T, Hembd A, Moore J, Montes D, Mosley T, Resendez J, *et al.* 2016. *TTG2* controls the developmental regulation of seed coat tannins in *Arabidopsis* by regulating vacuolar transport steps in the proanthocyanidin pathway. *Developmental Biology* **419**: 54–63.

Grant V. **1963**. *The origin of adaptations*. Columbia University Press, New York & London.

Grant V. **1981**. Factors promoting polyploidy. *Plant Speciation*: 307–323.

Guo M, Davis D, Birchler JA. **1996**. Dosage effects on gene expression in a maize ploidy series. *Genetics* **142**: 1349–1355.

Gutiérrez-Marcos JF, Dal Prà M, Giulini A, Costa LM, Gavazzi G, Cordelier S, Sellam O, Tatout C, Paul W, Perez P, *et al.* 2007. *empty pericarp4* encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant

growth in maize. *The Plant Cell* **19**: 196–210.

Gutierrez-Marcos JF, Pennington PD, Costa LM, Dickinson HG. **2003**. Imprinting in the endosperm: A possible role in preventing wide hybridization. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **358**: 1105–1111.

Haider N. **2013**. The origin of the B-genome of bread wheat (*Triticum aestivum* L.). *Russian Journal of Genetics* **49**: 263–274.

Haig D, Westoby M. **1991**. Genomic imprinting in endosperm: Its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philosophical Transactions: Biological Sciences* **333**: 1–13.

Hajjar R, Hodgkin T. **2007**. The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica* **156**: 1–13.

Hála M, Cole R, Synek L, Drdová E, Pecenková T, Nordheim A, Lamkemeyer T, Madlung J, Hochholdinger F, Fowler JE, *et al.* 2008. An exocyst complex functions in plant cell growth in *Arabidopsis* and Tobacco. *The Plant cell* **20**: 1330–1345.

Hao G-Y, Lucero ME, Sanderson SC, Zacharias EH, Holbrook NM. **2013**. Polyploidy enhances the occupation of heterogeneous environments through hydraulic related trade-offs in *Atriplex canescens* (Chenopodiaceae). *New Phytologist* **197**: 970–978.

Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA, Kay SA. **2000**. Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**: 2110–2113.

Hart GE. **1983**. Hexaploid wheat (*Triticum aestivum* L. em Thell). *Isozymes* **1**: 35–56.

Hassan L, Jones RN. **1994**. Long-range effects of colchicine sensitivity on meiosis in *Lolium multiflorum* L. (Italian ryegrass). *Heredity* **73**: 65–71.

Hassan L, Jones RN. **1995**. Colchicine-induced heritable modification of chiasma frequency and distribution in perennial ryegrass (*Lolium perenne*). *Genome* **38**: 356–361.

Hassan L, Jones RN, Parker JS, Posselt UK. **1991**. Colchicine-induced heritable variation in cell size and chloroplast number in the leaf cells of inbred ryegrasses (*Lolium perenne*, L. multiflorum). *Euphytica* **52**: 39–45.

Hegarty M, Coate J, Sherman-Broyles S, Abbott R, Hiscock S, Doyle J. **2013**. Lessons from natural and artificial polyploids in higher plants. *Cytogenetic and Genome Research* **140**: 204–225.

Hehenberger E, Kradolfer D, Kohler C, Köhler C. **2012**. Endosperm cellularization defines an important developmental transition for embryo development. *Development* **139**: 2031–2039.

Henderson IR, Salt DE. **2017**. Natural genetic variation and hybridization in plants. *Journal of Experimental Botany* **68**: 5415–5417.

Henry IM, Dilkes BP, Young K, Watson B, Wu H, Comai L. **2005**. Aneuploidy and genetic variation in the *Arabidopsis thaliana* triploid response. *Genetics* **170**: 1979–1988.

Heslop-Harrison JS, Schwarzacher T. **2007**. Domestication, genomics and the future for banana. *Annals of botany* **100**: 1073–1084.

Hias N, Leus L, Davey MW, Vanderzande S, Van Huylenbroeck J, Keulemans J. **2017**. Effect of polyploidization on morphology in two apple (*Malus × domestica*) genotypes. *Horticultural Science* **44**: 55–63.

Hias N, Svara A, Keulemans JW. **2018**. Effect of polyploidisation on the response of apple (*Malus × domestica* Borkh.) to *Venturia inaequalis* infection. *European Journal of Plant Pathology* **151**: 515–526.

Hu G, Wendel JF. **2019**. Cis–trans controls and regulatory novelty accompanying allopolyploidization. *New Phytologist* **221**: 1691–1700.

Hughes TE, Langdale JA, Kelly S. **2014**. The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome research* **24**: 1348–1355.

Hunt H V, Ansell SW, Russell SJ, Schneider H, Vogel JC. **2011**. Dynamics of polyploid formation and establishment in the allotetraploid rock fern *Asplenium majoricum*. *Annals of botany* **108**: 143–157.

Husband BC. **2004**. The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society* **82**: 537–546.

Husband BC, Baldwin SJ, Sabara HA. **2016**. Direct vs. indirect effects of whole-genome duplication on prezygotic isolation in *Chamerion angustifolium*: Implications for rapid speciation. *American Journal of Botany* **103**: 1259–1271.

Husband BC, Ozimec B, Martin SL, Pollock L. **2008**. Mating consequences of polyploid evolution in flowering plants: Current trends and insights from synthetic polyploids. *International Journal of Plant Sciences* **169**: 195–206.

Ishikawa R, Ohnishi T, Kinoshita Y, Eiguchi M, Kurata N, Kinoshita T. **2011**. Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *The Plant Journal* **65**: 798–806.

Iwasaki M, Paszkowski J. **2014**. Epigenetic memory in plants. *The EMBO journal* **33**: 1987–1998.

Jain A, Dashner ZS, Connolly EL. **2019**. Mitochondrial iron transporters (*MIT1* and *MIT2*) are essential for iron homeostasis and embryogenesis in *Arabidopsis thaliana*. *Frontiers in Plant Science* **10**: 1449.

Jeffrey Chen Z, Wang J, Tian L, Lee H-S, Wang JJ, Chen M, Lee JJ, Josefsson C, Madlung A, Watson B, *et al.* 2004. The development of an *Arabidopsis* model system for genome-wide analysis of polyploidy effects. *Biological journal of the Linnean Society. Linnean Society of London* **82**: 689–700.

Jenczewski E, Eber F, Grimaud A, Huet S, Lucas MO, Monod H, Chèvre AM. **2003**. *PrBn*, a major gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. *Genetics* **164**: 645–653.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, *et al.* 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97–100.

Johnston SA, Hanneman RE. **1982**. Manipulations of endosperm balance number overcome crossing barriers between diploid *Solanum* species. *Science* **217**: 446–8.

Josefsson C, Dilkes B, Comai L. **2006**. Parent-dependent loss of gene silencing during interspecies hybridization. *Current Biology* **16**: 1322–1328.

Joubès J, Chevalier C. **2000**. Endoreduplication in higher plants. *Plant Molecular Biology* **43**: 735–745.

Jullien PE, Berger F. **2010**. Parental genome dosage imbalance deregulates imprinting in *Arabidopsis*. *PLoS Genetics* **6**: e1000885.

Kaltenegger E, Leng S, Heyl A. **2018**. The effects of repeated whole genome duplication events on the evolution of cytokinin signaling pathway. *BMC Evolutionary Biology* **18**: 76.

Kamwean P, Chaisan T, Thobunluepop P, Phumichai C, Bredemeier M. **2017**. Changing of morphological characteristic and biomass properties in *Pennisetum purpureum* by colchicine treatment. *Journal of Agronomy* **16**: 23–31.

King KC, Seppälä O, Neiman M. **2012**. Is more better? Polyploidy and parasite resistance. *Biology Letters* **8**: 598–600.

Kirkbride RC, Yu HH, Nah G, Zhang C, Shi X, Chen ZJ. **2015**. An epigenetic role for disrupted paternal gene expression in postzygotic seed abortion in *Arabidopsis* interspecific hybrids. *Molecular Plant* **8**: 1766–1775.

Kitamura S, Akutsu M, Okazaki K. **2009**. Mechanism of action of nitrous oxide gas applied as a polyploidizing agent during meiosis in lilies. *Sexual Plant Reproduction* **22**: 9–14.

Klatt S, Schinkel CCF, Kirchheimer B, Dullinger S, Hörandl E. **2018**. Effects of cold treatments on fitness and mode of reproduction in the diploid and polyploid alpine plant *Ranunculus kuepferi* (Ranunculaceae). *Annals of botany* **121**: 1287–1298.

Knight CA, Beaulieu JM. **2008**. Genome size scaling through phenotype space. *Annals of botany* **101**: 759–766.

Köhler C, Mittelsten Scheid O, Erilova A. **2010**. The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in Genetics* **26**: 142–148.

Kondorosi E, Roudier F, Gendreau E. **2000**. Plant cell-size control: Growing by ploidy? *Current Opinion in Plant Biology* **3**: 488–492.

Koornneef M, Alonso-Blanco C, Vreugdenhil D. **2004**. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology* **55**: 141–172.

Kopecký D, Felder T, Schubiger FX, Mahelka V, Bartoš J, Doležel J, Boller B. **2018**. Frequent occurrence of triploid hybrids *Festuca pratensis × F. apennina* in the Swiss Alps. *Alpine Botany* **128**: 121–132.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R. **2009**. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics* **5**.

Kradolfer D, Wolff P, Jiang H, Siretskiy A, Köhler C. **2013**. An imprinted gene underlies postzygotic reproductive isolation in *Arabidopsis thaliana*. *Developmental Cell* **26**: 525–535.

Lafon-Placette C, Hatorangan MR, Steige KA, Cornille A, Lascoux M, Slotte T, Köhler C. **2018**. Paternally expressed imprinted genes associate with hybridization barriers in *Capsella*. *Nature Plants* **4**: 352–357.

Lafon-Placette C, Köhler C. **2016**. Endosperm-based postzygotic hybridization barriers: Developmental mechanisms and evolutionary drivers. *Molecular Ecology* **25**: 2620–2629.

Larkin JC, Walker JD, Bolognesi-Winfield AC, Gray JC, Walker AR. **1999**. Allele-specific interactions between *ttg* and *gl1* during trichome development in *Arabidopsis thaliana*. *Genetics* **151**: 1591–1604.

Lee HS, Chen ZJ. **2001**. Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 6753–6758.

Leitch IJ, Bennett MD. **1997**. Polyploidy in angiosperms. *Trends in Plant Science* **2**: 470–476.

Leitch IJ, Bennett MD. **2004**. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* **82**: 651–663.

Leitch IJ, Dodsworth S. **2017**. Endopolyploidy in plants. *eLS*: 1–10.

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. **2008**. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Annals of Botany* **101**: 805–814.

Leitch AR, Leitch IJ. **2008**. Genomic plasticity and the diversity of polyploid plants. *Science* **320**: 481–483.

Levin DA. **1983**. Polyploidy and novelty in flowering plants. *The American Naturalist* **122**: 1–25.

Li M, Ding B, Huang W, Pan J, Ding Z, Jiang F. **2018**. Induction and characterization of tetraploids from seeds of *Bletilla striata* (Thunb.) Reichb.f. *BioMed Research International* **2018**: 3246398.

Li J, Hou R, Niu X, Liu R, Wang Q, Wang C, Li X, Hao Z, Yin G, Zhang K. **2016**. Comparison of microarray and RNA-Seq analysis of mRNA expression in dermal mesenchymal stem cells. *Biotechnology Letters* **38**: 33–41.

Li S, Liu L, Li T, Lan T, Wang Y, Zhang Z, Liu J, Xu S, Zhang X, Zhu J, *et al.* 2019. The distribution pattern of endopolyploidy in maize. *Theoretical and Applied Genetics* **132**: 1487–1503.

Lignowski EM, Scott EGG. 1972. Effect of trifluralin on mitosis. *Weed Science* **20**: 267–270.

Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR. **2007**. Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytologist* **175**: 756–763.

Lin BY. **1984**. Ploidy barrier to endosperm development in maize. *Genetics* **107**: 103–15.

Lin Y, Schiefelbein J. **2001**. Embryonic control of epidermal cell patterning in the root and hypocotyl of *Arabidopsis*. *Development* **128**: 3697–3705.

Liqin G, Jianguo Z, Xiaoxia L, Guodong R. **2019**. Polyploidy-related differential gene expression between diploid and synthesized allotriploid and allotetraploid hybrids of *Populus*. *Molecular Breeding* **39**: 69.

Liu S-L, Adams KL. **2010**. Dramatic change in function and expression pattern of a gene duplicated by polyploidy created a paternal effect gene in the Brassicaceae. *Molecular Biology and Evolution* **27**: 2817–2828.

Lou P, Wu J, Cheng F, Cressman LG, Wang X, McClung CR. **2012**. Preferential retention of circadian clock genes during diploidization following whole genome triplication in *Brassica rapa*. *The Plant Cell* **24**: 2415–2426.

Lu J, Zhang C, Baulcombe DC, Chen ZJ. **2012**. Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 5529–5534.

Luckett DJ. **1989**. Colchicine mutagenesis is associated with substantial heritable variation in cotton. *Euphytica* **42**: 177–182.

Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T. **2006**. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant physiology* **140**: 336–348.

Lynch M, Conery JS. **2000**. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.

Ma J, Zhang H, Li S, Zou Y, Li T, Liu J, Ding P, Mu Y, Tang H, Deng M, *et al.* 2019. Identification of quantitative trait loci for kernel traits in a wheat cultivar Chuannong16. *BMC Genetics* **20**: 77.

Mable BK, Alexandrou MA, Taylor MI. **2011**. Genome duplication in amphibians and

fish: An extended synthesis. *Journal of Zoology* **284**: 151–182.

Madlung A. **2013**. Polyploidy and its effect on evolutionary success: Old questions revisited with new tools. *Heredity* **110**: 99–104.

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. **2002**. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant physiology* **129**: 733–746.

Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L. **2005**. Genomic changes in synthetic *Arabidopsis* polyploids. *The Plant Journal* **41**: 221–230.

Maluszynski M., Kasha K.J., Forster B.P., Szarejko I., eds. 2003. Cytogenetic tests for ploidy level analyses — chromosome counting. In: *Doubled Haploid Production in Crop Plants: A manual*. Dordrecht, Netherlands: Kluwer Academic Publishers. 391–395.

Mandáková T, Li Z, Barker MS, Lysak MA. **2017**. Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *The Plant Journal* **91**: 3–21.

Mandáková T, Lysak MA. **2018**. Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology* **42**: 55–65.

Marfil CF, Duarte PF, Masuelli RW. **2018**. Phenotypic and epigenetic variation induced in newly synthesized allopolyploids and autopolyploids of potato. *Scientia Horticulturae* **234**: 101–109.

Markov A V, Kaznacheev IS. **2016**. Evolutionary consequences of polyploidy in prokaryotes and the origin of mitosis and meiosis. *Biology direct* **11**: 28.

Marks GE. **1966**. The enigma of triploid potatoes. *Euphytica* **15**: 285–290.

Martelotto LG, Ortiz JPA, Stein J, Espinoza F, Quarin CL, Pessino SC. **2007**. Genome rearrangements derived from autopolyploidization in *Paspalum* sp. *Plant Science* **172**: 970–977.

Masterson J. **1994**. Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. *Science* **264**: 421–424.

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. **2011**. Recently formed polyploid plants diversify at lower rates. *Science* **333**: 1257–1257.

McCarthy DJ, Chen Y, Smyth GK. **2012**. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**: 4288–4297.

McIntyre PJ, Strauss S. **2017**. An experimental test of local adaptation among cytotypes within a polyploid complex. *Evolution* **71**: 1960–1969.

Mena M, Ambrose BA, Meeley RB, Briggs SP, Yanofsky MF, Schmidt RJ. **1996**. Diversification of C-function activity in maize flower development. *Science* **274**: 1537–

1540.

Mendell JE, Clements KD, Choat JH, Angert ER. **2008**. Extreme polyploidy in a large bacterium. *Proceedings of the National Academy of Sciences* **105**: 6730–6734.

Meng H bing, Jiang S si, Hua S jin, Lin X yong, Li Y long, Guo W li, Jiang L xi. **2011**. Comparison between a tetraploid turnip and its diploid progenitor (*Brassica rapa* L.): The adaptation to salinity stress. *Agricultural Sciences in China* **10**: 363–375.

Mitchell-Olds T, Willis JH, Goldstein DB. **2007**. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics* **8**: 845–856.

Monda K, Araki H, Kuhara S, Ishigaki G, Akashi R, Negi J, Kojima M, Sakakibara H, Takahashi S, Hashimoto-Sugimoto M, *et al.* 2016. Enhanced stomatal conductance by a spontaneous *Arabidopsis* tetraploid, Me-0, results from increased stomatal size and greater stomatal aperture. *Plant Physiology* **170**: pp.01450.2015.

Moore RC, Purugganan MD. **2005**. The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology* **8**: 122–128.

Münzbergová Z. **2017**. Colchicine application significantly affects plant performance in the second generation of synthetic polyploids and its effects vary between populations. *Annals of Botany* **120**: 329–339.

Münzbergová Z, Skuhrovec J. **2017**. Contrasting effects of ploidy level on seed production in a diploid–tetraploid system. *AoB PLANTS* **9**: plw077.

Nagl W. **1976**. DNA endoreduplication and polyteny understood as evolutionary strategies. *Nature* **261**: 614–615.

Nebel BR. **1937**. Mechanism of polyploidy through colchicine. *Nature* **140**: 1101.

Nellist CF, Vickerstaff RJ, Sobczyk MK, Marina-Montes C, Wilson FM, Simpson DW, Whitehouse AB, Harrison RJ. **2019**. Quantitative Trait Loci controlling *Phytophthora cactorum* resistance in the cultivated octoploid strawberry (*Fragaria × ananassa*). *Horticulture Research* **6**: 60.

De Nettancourt D, Dijkhuis P, van Gastel AJGG, Broertjes C. **1971**. The combined use of leaf irradiation and of the adventitious bud technique for inducing and detecting polyploidy, marker mutations and self-compatibility in clonal populations of *Nicotiana alata* LINK and OTTO. *Euphytica* **20**: 508–520.

Ni Z, Kim E-D, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ. **2009**. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* **457**: 327–331.

Nicolas SD, Monod H, Eber F, Chèvre A-M, Jenczewski E. **2012**. Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization. *The Plant Journal* **70**: 691–703.

Ochatt SJ. **2008**. Flow cytometry in plant breeding. *Cytometry Part A* **73A**: 581–598.

Osborn TC, Chris Pires J, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A,

Doerge RW, Colot V, *et al.* 2003. Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics* **19**: 141–147.

Oswald BP, Nuismer SL. **2011**. A unified model of autopolyploid establishment and evolution. *The American Naturalist* **178**: 687–700.

Otto SP, Whitton J. **2000**. Polyploid incidence and evolution. *Annual Review of Genetics* **34**: 401–437.

Palanivelu R, Tsukamoto T. **2012**. Pathfinding in angiosperm reproduction: pollen tube guidance by pistils ensures successful double fertilization. *Wiley interdisciplinary reviews. Developmental biology* **1**: 96–113.

Parisod C, Holderegger R, Brochmann C. **2010**. Evolutionary consequences of autopolyploidy. *The New Phytologist* **186**: 5–17.

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien M-A, Ainouche M. **2009**. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytologist* **184**: 1003–1015.

Parks MB, Nakov T, Ruck EC, Wickett NJ, Alverson AJ. **2018**. Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). *American Journal of Botany* **105**: 330–347.

Paterson A. **2008**. Paleopolyploidy and its impact on the structure and function of modern plant genomes. In: *Genome Dynamics*. Basel: Karger. **4**: 1–12.

Van de Peer Y, Mizrachi E, Marchal K. **2017**. The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18**: 411–424.

Pellicer J, Clermont S, Houston L, Rich TCG, Fay MF. **2012**. Cytotype diversity in the *Sorbus* complex (Rosaceae) in Britain: Sorting out the puzzle. *Annals of Botany* **110**: 1185–1193.

Piferrer F, Beaumont A, Falguière J-C, Flajšhans M, Haffray P, Colombo L. **2009**. Polyploid fish and shellfish: Production, biology and applications to aquaculture for performance improvement and genetic containment. *Aquaculture* **293**: 125–156.

Pignatta D, Dilkes BP, Yoo SY, Henry IM, Madlung A, Doerge RW, Jeffrey Chen Z, Comai L. **2010**. Differential sensitivity of the *Arabidopsis thaliana* transcriptome and enhancers to the effects of genome doubling. *New Phytologist* **186**: 194–206.

Pignocchi C, Minns GE, Nesi N, Koumproglou R, Kitsios G, Benning C, Lloyd CW, Doonan JH, Hills MJ. **2009**. *ENDOSPERM DEFECTIVE1* is a novel microtubule-associated protein essential for seed development in *Arabidopsis*. *The Plant Cell* **21**: 90–105.

Pine PS, Munro SA, Parsons JR, McDaniel J, Lucas AB, Lozach J, Myers TG, Su Q, Jacobs-Helber SM, Salit M. **2016**. Evaluation of the External RNA Controls Consortium (ERCC) reference material using a modified Latin square design. *BMC biotechnology* **16**: 54.

Pires JC, Zhao J, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC. **2004**. Flowering time divergence and genomic rearrangements in resynthesized *Brassica*

polyploids (Brassicaceae). *Biological Journal of the Linnean Society* **82**: 675–688.

Planchais S, Glab N, Inzé D, Bergounioux C. **2000**. Chemical inhibitors: a tool for plant cell cycle studies. *FEBS Letters* **476**: 78–83.

Poggio L, González GE. **2018**. Cytological diploidization of paleopolyploid genus *Zea*: Divergence between homoeologous chromosomes or activity of pairing regulator genes? *PloS one* **13**: e0189644–e0189644.

Prohens J, Gramazio P, Plazas M, Dempewolf H, Kilian B, Díez MJ, Fita A, Herraiz FJ, Rodríguez-Burruezo A, Soler S, *et al.* 2017. Introgressiomics: A new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica* **213**: 158.

Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, Browse J, Chapple C, Colot V, Cutler S, *et al.* 2016. 50 years of *Arabidopsis* research: Highlights and future directions. *New Phytologist* **209**: 921–944.

De Queiroz K. **2007**. Species concepts and species delimitation. *Systematic Biology* **56**: 879–886.

Ramsey J. **2011**. Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences* **108**: 7096–7101.

Ramsey J, Schemske DW. **1998**. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**: 467–501.

Ramsey J, Schemske DW. **2002**. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* **33**: 589–639.

Rastogi S, Liberles DA. **2005**. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evolutionary Biology* **5**: 28.

Rebernig CA, Lafon-Placette C, Hatorangan MR, Slotte T, Köhler C. **2015**. Non-reciprocal interspecies hybridization barriers in the *Capsella* genus are established in the endosperm. *PLoS Genetics* **11**: 1–19.

Renny-Byfield S, Kovarik A, Kelly LJ, Macas J, Novak P, Chase MW, Nichols RA, Pancholi MR, Grandbastien M-A, Leitch AR. **2013**. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *The Plant Journal* **74**: 829–839.

Renny-Byfield S, Wendel JF. **2014**. Doubling down on genomes: Polyploidy and crop plants. *American Journal of Botany* **101**: 1711–1725.

Riddle NC, Jiang H, An L, Doerge RW, Birchler JA. **2010**. Gene expression analysis at the intersection of ploidy and hybridity in maize. *Theoretical and Applied Genetics* **120**: 341–353.

Riddle NC, Kato A, Birchler JA. **2006**. Genetic variation for the response to ploidy change in *Zea mays* L. *Theoretical and Applied Genetics* **114**: 101–111.

Risso D, Schwartz K, Sherlock G, Dudoit S. **2011**. GC-content normalization for RNA-Seq

data. *BMC bioinformatics* **12**: 480.

Robinson DO, Coate JE, Singh A, Hong L, Bush MS, Doyle JJ, Roeder AHK, Jeff J, Doyle JJ, Roeder AHK. **2018**. Ploidy and size at multiple scales in the *Arabidopsis* sepal. *The Plant Cell* **30**: 2308–2329.

Robinson MD, McCarthy DJ, Smyth GK. **2010**. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Roeder AHK, Chickarmane V, Cunha A, Obara B, Manjunath BS, Meyerowitz EM. **2010**. Variability in the control of cell division underlies sepal epidermal patterning in *Arabidopsis thaliana*. *PLoS biology* **8**: e1000367–e1000367.

Roth M, Florez-Rueda AM, Städler T. **2019**. Differences in effective ploidy drive genome-wide endosperm expression polarization and seed failure in wild tomato hybrids. *Genetics* **212**: 141–152.

Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA. **2013**. The fate of duplicated genes in a polyploid plant genome. *The Plant Journal* **73**: 143–153.

Salmon A, Ainouche ML, Wendel JF. **2005**. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* **14**: 1163–1175.

Saminathan T, Nimmakayala P, Manohar S, Malkaram S, Almeida A, Cantrell R, Tomason Y, Abburi L, Rahman MA, Vajja VG, *et al.* 2015. Differential gene expression and alternative splicing between diploid and tetraploid watermelon. *Journal of experimental botany* **66**: 1369–1385.

Santos JL, Alfaro D, Sanchez-Moran E, Armstrong SJ, Franklin FCH, Jones GH. **2003**. Partial diploidization of meiosis in autotetraploid *Arabidopsis thaliana*. *Genetics* **165**: 1533–1540.

Sattler MC, Carvalho CR, Clarindo WR. **2016**. The polyploidy and its key role in plant breeding. *Planta* **243**: 281–296.

Scarcelli N, Cheverud JM, Schaal BA, Kover PX. **2007**. Antagonistic pleiotropic effects reduce the potential adaptive value of the FRIGIDA locus. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 16986–91.

Schatlowski N, Kohler C. **2012**. Tearing down barriers: Understanding the molecular mechanisms of interploidy hybridizations. *Journal of Experimental Botany* **63**: 6059–6067.

Schatlowski N, Wolff P, Santos-Gonzalez J, Schoft V, Siretskiy A, Scott R, Tamaru H, Kohler C. **2014a**. Hypomethylated pollen bypasses the interploidy hybridization barrier in *Arabidopsis*. *The Plant Cell* **26**: 3556–3568.

Schatlowski N, Wolff P, Santos-González J, Schoft V, Siretskiy A, Scott R, Tamaru H, Köhler C. **2014b**. Hypomethylated pollen bypasses the interploidy hybridization barrier

in *Arabidopsis*. *The Plant Cell* **26**: 3556–3568.

Schinkel CCF, Kirchheimer B, Dullinger S, Geelen D, De Storme N, Hörandl E. **2017**. Pathways to polyploidy: Indications of a female triploid bridge in the alpine species *Ranunculus kuepferi* (Ranunculaceae). *Plant Systematics and Evolution* **303**: 1093–1108.

Schneider CA, Rasband WS, Eliceiri KW. **2012**. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671.

Scott RJ, Spielman M, Bailey J, Dickinson HG, Abbott RJ, Gomes MF, Bhatnagar SP, Sawhney V, Birchler JA, Braselton JP, *et al.* 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* **125**: 3329–3341.

Sekine D, Ohnishi T, Furuumi H, Ono A, Yamada T, Kurata N, Kinoshita T. **2013**. Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *The Plant Journal* **76**: 792–799.

Semeniuk P, Arisumi T. **1968**. Colchicine-induced tetraploid and cytochimeral roses. *Botanical Gazette* **129**: 190–193.

Seoighe C, Gehring C. **2004**. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends in Genetics* **20**: 461–464.

Serapiglia MJ, Gouker FE, Smart LB. **2014**. Early selection of novel triploid hybrids of shrub willow with improved biomass yield relative to diploids. *BMC plant biology* **14**: 74.

Shi X, Zhang C, Ko DK, Jeffrey Chen Z. **2015**. Genome-wide dosage-dependent and-independent regulation contributes to gene expression and evolutionary novelty in plant polyploids. *Molecular Biology and Evolution* **32**: 2351–2366.

Shimizu-Inatsugi R, Terada A, Hirose K, Kudoh H, Sese J, Shimizu KK. **2017**. Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Molecular Ecology* **26**: 193–207.

Shu Z, Row S, Deng W-M. **2018**. Endoreplication: The good, the bad, and the ugly. *Trends in cell biology* **28**: 465–474.

Símová I, Herben T. **2012**. Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. *Proceedings. Biological sciences* **279**: 867–875.

De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y. **2013**. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences* **110**: 2898–2903.

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev E V, Mei W, Cortez MB, Soltis PS, Gitzendanner MA. **2014**. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose *et al*. (2011). *New Phytologist* **202**: 1105–1117.

Soltis DE, Soltis PS, Tate JA. **2004**. Advances in the study of polyploidy since plant

speciation. *New Phytologist* **161**: 173–191.

Soneson C, Love MI, Robinson MD. **2015**. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research* **4**: 1521.

Song Q, Ando A, Xu D, Fang L, Zhang T, Huq E, Qiao H, Deng XW, Chen ZJ. **2018**. Diurnal down-regulation of ethylene biosynthesis mediates biomass heterosis. *Proceedings of the National Academy of Sciences* **115**: 5606–5611.

Song Q, Chen ZJ. **2015**. Epigenetic and developmental regulation in plant polyploids. *Current opinion in plant biology* **24**: 101–109.

Soppa J. **2014**. Polyploidy in Archaea and Bacteria: About desiccation resistance, giant cell size, long-term survival, enforcement by a eukaryotic host and additional aspects. *Journal of Molecular Microbiology and Biotechnology* **24**: 409–419.

Sorensen MB, Mayer U, Lukowitz W, Robert H, Chambrier P, Jurgens G, Somerville C, Lepiniec L, Berger F. **2002**. Cellularisation in the endosperm of *Arabidopsis thaliana* is coupled to mitosis and shares multiple components with cytokinesis. *Development* **129**: 5567–5576.

Spoelhof JP, Soltis PS, Soltis DE. **2017**. Pure polyploidy: Closing the gaps in autopolyploid research. *Journal of Systematics and Evolution* **55**: 340–352.

Stebbins GL. **1971**. *Chromosomal evolution in higher plants*. London: Edward Arnold.

Stebbins Jr. CL. **1950**. *Variation and evolution in plants.* London: Oxford University Press.

De Storme N, Copenhaver GP, Geelen D. **2012**. Production of diploid male gametes in *Arabidopsis* by cold-induced destabilization of postmeiotic radial microtubule arrays. *Plant Physiology* **160**: 1808–1826.

De Storme N, Mason A. **2014**. Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance. *Current Plant Biology* **1**: 10–33.

Stoute AI, Varenko V, King GJ, Scott RJ, Kurup S. **2012**. Parental genome imbalance in *Brassica oleracea* causes asymmetric triploid block. *The Plant Journal* **71**: 503–516.

Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyang S, Veilleux RE, Busse JS, Erhardt RJ, Buell CR, *et al.* 2007. Phenotypic and transcriptomic changes associated with potato autopolyploidization. *Genetics* **176**: 2055–2067.

Subramaniam S, Wang X, Freeling M, Pires JC. **2013**. The fate of *Arabidopsis thaliana* homeologous CNSs and their motifs in the paleohexaploid *Brassica rapa*. *Genome Biology and Evolution* **5**: 646–660.

Sugimoto-Shirasu K, Stacey NJ, Corsar J, Roberts K, McCann MC. **2002**. DNA topoisomerase VI is essential for endoreduplication in *Arabidopsis*. *Current Biology* **12**: 1782–1786.

Sukno S, Ruso J, Jan CC, Melero-Vara JM, Fernández-martínez JM. **1999**. Interspecific

hybridization between sunflower and wild perennial *Helianthus* species via embryo rescue. *Euphytica* **106**: 69–78.

Sweigart AL, Martin NH, Willis JH. **2008**. Patterns of nucleotide variation and reproductive isolation between a *Mimulus* allotetraploid and its progenitor species. *Molecular Ecology* **17**: 2089–2100.

Szadkowski E, Eber F, Huteau V, Lodé M, Coriton O, Jenczewski E, Chèvre AM. **2011**. Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. *New Phytologist* **191**: 884–894.

Tang Z-Q, Chen D-L, Song Z-J, He Y-C, Cai D-T. **2010**. In vitro induction and identification of tetraploid plants of *Paulownia tomentosa*. *Plant Cell, Tissue and Organ Culture (PCTOC)* **102**: 213–220.

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC. **2012**. Altered patterns of fractionation and exon deletions in *Brassica* rapa support a two-step model of paleohexaploidy. *Genetics* **190**: 1563–1574.

Tayalé A, Parisod C. **2013**. Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenetic and Genome Research* **140**: 79–96.

Thiebaut F, Hemerly AS, Ferreira PCG. **2019**. A role for epigenetic regulation in the adaptation and stress responses of non-model plants. *Frontiers in plant science* **10**: 246.

Thomas BC, Pedersen B, Freeling M. **2006**. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome research* **16**: 934–946.

Thornber CS. **2006**. Functional properties of the isomorphic biphasic algal life cycle. *Integrative and Comparative Biology* **46**: 605–614.

Tian L, Li X, Ha M, Zhang C, Chen ZJ. **2014**. Genetic and epigenetic changes in a genomic region containing MIR172 in *Arabidopsis* allopolyploids and their progenitors. *Heredity* **112**: 207–214.

Todd RT, Forche A, Selmecki A. **2017**. Ploidy variation in fungi: Polyploidy, aneuploidy, and genome evolution. *Microbiology spectrum* **5**: 10.1128/microbiolspec.FUNK-0051–2016.

Tsukaya H. **2008**. Controlling size in multicellular organs: Focus on the leaf. *PLoS Biology* **6**: 1373–1376.

Tsukaya H. **2013**. Does ploidy level directly control cell size? Counterevidence from *Arabidopsis* genetics. *PLoS ONE* **8**: 1–7.

Udall JA, Wendel JF. **2006**. Polyploidy and crop improvement. *Crop Science* **46**: 3-14.

Valentine DH, Woodell SRJ. **1963**. Studies in British Primulas. *New Phytologist* **62**: 125–143.

Vallejo-Marín M, Hiscock SJ. **2016**. Hybridization and hybrid speciation under global

change. *New Phytologist* **211**: 1170–1187.

Vazzola V, Losa A, Soave C, Murgia I. **2007**. Knockout of frataxin gene causes embryo lethality in *Arabidopsis*. *FEBS Letters* **581**: 667–672.

Vogt G. **2017**. Facilitation of environmental adaptation and evolution by epigenetic phenotype variation: Insights from clonal, invasive, polyploid, and domesticated animals. *Environmental Epigenetics* **3**.

Wagner Jr. WH. **1970**. Biosystematics and evolutionary noise. *TAXON* **19**: 146–151.

Wang J. **2004**. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* **167**: 1961–1973.

Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, *et al.* 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* **172**: 507–517.

Wang M, Tu L, Yuan D, Zhu D, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, *et al.* 2019. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nature Genetics* **51**: 224–229.

Weeks AR, Marec F, Breeuwer JAJ. **2001**. A mite species that consists entirely of haploid females. *Science* **292**: 2479–2482.

Wei S, Weiss ZR, Williams Z. **2018**. Rapid multiplex small DNA sequencing on the MinION nanopore sequencing platform. *G3: Genes, Genomes, Genetics* **8**: 1649–1657.

Weigel D. **2012**. Natural variation in Arabidopsis: From molecular genetics to ecological genomics. *Plant Physiology* **158**: 2–22.

Weiss H, Maluszynska J. **2000**. Chromosomal rearrangement in autotetraploid plants of *Arabidopsis thaliana*. *Hereditas* **133**: 255–261.

Welles SR, Ellstrand NC. **2019**. Evolution of increased vigour associated with allopolyploidization in the newly formed invasive species *Salsola ryanii*. *AoB PLANTS* **12**: plz039.

Wilhelm BT, Landry JR. **2009**. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**: 249–257.

Winkler H. **1917**. Über die experimentelle Erzeugung yon Pflanzen mit abweichenden Chromosomenzahlen. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* **17**: 270–272.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. **2009**. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 13875–13879.

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. **2010**. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS biology* **8**: e1000409–e1000409.

Wright KM, Pires JC, Madlung A. **2009**. Mitotic instability in resynthesized and natural polyploids of the genus *Arabidopsis* (Brassicaceae). *American Journal of Botany* **96**: 1656–1664.

Xiang D, Quilichini TD, Liu Z, Gao P, Pan Y, Li Q, Nilsen KT, Venglat P, Esteban E, Pasha A, *et al.* 2019. The transcriptional landscape of polyploid wheats and their diploid ancestors during embryogenesis and grain development. *The Plant Cell* **31**: 2888–2911.

Xu Y, Zhong L, Wu X, Fang X, Wang J. **2009**. Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta* **229**: 471–483.

Yang J, Yang G, Yang M, Su L, Xia A, Li D, Huang C, Zhou D, Liu Y, Wang H, *et al.* 2019. Quantitative Trait Locus analysis of seed germination and early seedling growth in rice. *Frontiers in Plant Science* **10**: 1582.

Yant L, Bomblies K. **2015**. Genome management and mismanagement—cell-level opportunities and challenges of whole-genome duplication. *Genes & Development* **29**: 2405–2419.

Yao H, Kato A, Mooney B, Birchler JA. **2011**. Phenotypic and gene expression analyses of a ploidy series of maize inbred Oh43. *Plant Molecular Biology* **75**: 237–251.

Yin L, Qu J, Zhou H, Shang X, Fang H, Lu J, Yan H. **2018**. Comparison of leaf transcriptomes of cassava "Xinxuan 048" diploid and autotetraploid plants. *Genes & Genomics* **40**: 927–935.

Yoo M-J, Szadkowski E, Wendel JF. **2013**. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**: 171–180.

Yu Z, Haage K, Streit VE, Gierl A, Torres Ruiz RA. **2009**. A large number of tetraploid *Arabidopsis thaliana* lines, generated by a rapid strategy, reveal high stability of neo-tetraploids during consecutive generations. *Theoretical and Applied Genetics* **118**: 1107–1119.

Yu Z, Haberer G, Matthes M, Rattei T, Mayer KFXX, Gierl A, Torres-Ruiz RA. **2010**. Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 17809–17814.

Zhang N, Bao Y, Xie Z, Huang X, Sun Y, Feng G, Zeng H, Ren J, Li Y, Xiong J, *et al.* 2019a. Efficient Characterization of Tetraploid Watermelon. *Plants (Basel, Switzerland)* **8**: 419.

Zhang XY, Hu CG, Yao JL. **2010**. Tetraploidization of diploid *Dioscorea* results in activation of the antioxidant defense system and increased heat tolerance. *Journal of Plant Physiology* **167**: 88–94.

Zhang T, Yu L-X, Zheng P, Li Y, Rivera M, Main D, Greene SL. **2015**. Identification of loci associated with drought resistance traits in heterozygous autotetraploid Alfalfa (*Medicago sativa* L.) using genome-wide association studies with genotyping by sequencing. *PloS one* **10**: e0138931–e0138931.

Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, *et al.* 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics* **50**: 1565–1573.

Zhang H, Zheng R, Wang Y, Zhang Y, Hong P, Fang Y, Li G, Fang Y. **2019b**. The effects of *Arabidopsis* genome duplication on the chromatin organization and transcriptional regulation. *Nucleic Acids Research* **47**: 7857–7869.

Zhou R, Moshgabadi N, Adams KL. **2011**. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proceedings of the National Academy of Sciences* **108**: 16122–16127.

Supplementary Data 7.1 | Primer pairs used for genotyping. Forward and reverse primer sequences are provided for five primer pair. Each pair corresponds to a large indel on each of the five chromosomes. The expected PCR product length for each genotype with each primer pair is provided.

Primer Sequence (5' - 3')	Type	Chromosome:indel location	Expected PCR product length (in basepairs)						
			Col	Ct	Kn	Ler	Mt	No	Wu
GTTCAGGATTTCAGGAGGTA	Forward	Chr1:indel3196228	447	447	256	256	256	256	447
TTGGAAGCTTGTATCCATCC	Reverse								
TGTTTTGGTGAGGGTTGACAAA	Forward	Chr2:indel5407623	485	582	485	528	485	528	485
AGCCCGGATTTTCTCCTAGT	Reverse								
TACTACCAGAGATGCAAACG	Forward	Chr3:indel11672524	673	495	495	673	673	495	495
GGATACTTCAAAACCATTATGATAC	Reverse								
GCAGTGGTGATGAAGAAAAG	Forward	Chr4:indel13960433	411	296	296	296	411	296	411
TAGAGATGTGGGACAAGTGA	Reverse								
TCGTCTTCTACACCAGATCA	Forward	Chr5:indel8219475	490	490	490	354	354	490	490
ATGCATGGAACATAACCTCT	Reverse								

Supplementary Data 7.2 | Summary statistics table for colchicine treated diploids and tetraploids from seven genotypes for ten phenotypic traits. SD = standard deviation. CV = coefficient of variance.

Trait	Genotype	$2x_{\text{-colchicine}}$			$4x_{\text{-colchicine}}$		
		Mean	SD	CV	Mean	SD	CV
Flowering Time (Days)	Col	29.367	3.873	13.188	27.500	2.030	7.382
	Ct	24.947	2.297	9.206	26.525	1.867	7.040
	Kn	35.069	3.093	8.820	37.963	3.492	9.197
	Ler	23.310	1.713	7.351	25.346	3.509	13.846
	Mt	31.400	2.943	9.373	31.600	2.268	7.178
	No	29.900	3.210	10.736	30.100	2.580	8.571
	Wu	27.867	2.403	8.623	27.000	1.338	4.954
Rosette Diameter (cm)	Col	3.330	0.584	17.531	3.703	0.410	11.065
	Ct	3.753	1.185	31.585	4.608	0.546	11.841
	Kn	3.479	0.827	23.755	3.940	0.647	16.430
	Ler	3.079	0.678	22.015	2.704	0.589	21.780
	Mt	3.245	0.560	17.249	3.175	0.577	18.177
	No	2.572	0.762	29.632	3.578	0.660	18.439
	Wu	3.317	0.484	14.602	4.055	0.504	12.433
Number of Leaves (15 days post germination)	Col	12.500	1.253	10.021	11.700	0.837	7.151
	Ct	11.789	1.475	12.511	12.450	0.959	7.706
	Kn	14.276	1.360	9.527	12.900	1.423	11.029
	Ler	11.172	1.136	10.168	9.462	1.067	11.277
	Mt	13.172	1.284	9.745	12.000	1.247	10.393
	No	10.444	1.423	13.629	10.800	1.181	10.936
	Wu	14.000	1.114	7.958	14.100	1.683	11.934
Leaf Area (cm2)	Col	2.322	0.631	27.171	2.279	0.372	16.330
	Ct	2.621	0.437	16.682	2.793	1.081	38.714
	Kn	2.101	0.530	25.243	3.092	0.743	24.033
	Ler	2.584	1.128	43.653	2.919	0.698	23.906
	Mt	2.192	0.651	29.676	2.508	0.593	23.662
	No	2.639	0.712	26.961	3.447	0.780	22.612
	Wu	2.092	0.428	20.467	2.932	0.702	23.940
Trichome Density (trichomes per cm2)	Col	64.399	15.019	23.321	64.821	15.693	24.210
	Ct	48.261	15.122	31.333	45.932	19.080	41.538
	Kn	29.957	7.113	23.743	23.643	13.162	55.670
	Ler	35.618	15.091	42.370	26.589	11.914	44.807
	Mt	88.091	15.806	17.942	74.533	22.771	30.551
	No	33.812	20.067	59.349	25.332	7.254	28.637
	Wu	63.156	13.107	20.754	78.461	27.923	35.588

Trait	Genotype	2x-colchicine			4x-colchicine		
		Mean	SD	CV	Mean	SD	CV
Total Trichomes (per leaf)	Col	143.414	37.349	26.043	144.655	33.120	22.896
	Ct	125.579	41.236	32.836	121.152	36.745	30.330
	Kn	61.038	17.280	28.310	67.708	16.931	25.005
	Ler	84.964	33.756	39.729	75.000	26.831	35.775
	Mt	189.567	51.614	27.228	182.036	52.527	28.855
	No	81.350	28.463	34.988	83.629	16.821	20.114
	Wu	130.800	33.458	25.579	227.526	82.004	36.042
Trichomes with >3 branches (% of total trichomes on the leaf)	Col	27.450	5.781	21.061	62.455	21.277	34.068
	Ct	21.312	8.152	38.249	61.108	18.419	30.142
	Kn	15.382	6.997	45.488	83.986	8.548	10.177
	Ler	19.880	7.622	38.339	73.896	17.641	23.872
	Mt	0.745	0.692	92.945	19.137	14.029	73.304
	No	2.164	1.668	77.103	45.152	9.478	20.991
	Wu	30.444	6.505	21.366	82.253	12.441	15.125
Stomatal Density (per mm2)	Col	130.904	29.688	22.679	65.923	14.784	22.426
	Ct	102.416	58.679	57.295	65.805	12.862	19.545
	Kn	146.669	40.877	27.870	63.097	14.818	23.485
	Ler	172.879	31.099	17.989	89.466	17.758	19.849
	Mt	171.870	38.329	22.301	95.016	18.126	19.077
	No	156.174	65.172	41.730	84.051	18.455	21.957
	Wu	157.587	19.463	12.351	84.051	12.248	14.572
Stomatal Diameter (µm)	Col	21.797	2.490	11.423	33.159	3.089	9.315
	Ct	22.432	3.362	14.987	30.424	2.558	8.407
	Kn	22.155	2.684	12.113	32.777	2.889	8.816
	Ler	22.916	2.025	8.839	32.182	3.168	9.843
	Mt	22.628	2.231	9.860	31.217	3.692	11.826
	No	19.818	2.530	12.767	31.383	2.848	9.074
	Wu	22.355	2.157	9.647	35.178	3.891	11.061
Total Fruits	Col	249.700	78.246	31.336	139.586	43.644	31.267
	Ct	194.053	51.646	26.614	217.615	87.520	40.218
	Kn	311.300	65.920	21.176	245.600	51.868	21.119
	Ler	301.517	96.307	31.941	357.962	102.079	28.517
	Mt	258.067	71.437	27.682	300.733	94.737	31.502
	No	326.000	100.919	30.957	248.425	72.491	29.180
	Wu	331.067	96.146	29.041	305.800	95.102	31.100

Supplementary Data 7.3 | Output table of Tukey's HSD test for ANOVAs that tested significant for the effects of line within genotype-by-ploidy ((Genotype x Ploidy)/Line). Tukey comparisons within lines of the same genotype have been presented here. Lower and upper bound represent the lower and upper limits of the 95% confidence interval. Sig represents the adjusted P-value for the comparison corrected for multiple comparisons.

Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
	COL 2x - B	COL 2x - A	1.800	-4.330	7.930	1.000
	COL 2x - C	COL 2x - A	4.100	-2.030	10.230	0.999
	COL 2x - C	COL 2x - B	2.300	-3.830	8.430	1.000
	COL 4x - B	COL 4x - A	-0.600	-6.730	5.530	1.000
	COL 4x - C	COL 4x - A	0.600	-5.530	6.730	1.000
	COL 4x - C	COL 4x - B	1.200	-4.930	7.330	1.000
	CT 2x - C	CT 2x - B	0.311	-5.987	6.609	1.000
	CT 4x - B	CT 4x - A	0.500	-5.630	6.630	1.000
	CT 4x - C	CT 4x - A	-1.000	-7.130	5.130	1.000
	CT 4x - D	CT 4x - A	0.600	-5.530	6.730	1.000
	CT 4x - C	CT 4x - B	-1.500	-7.630	4.630	1.000
	CT 4x - D	CT 4x - B	0.100	-6.030	6.230	1.000
	CT 4x - D	CT 4x - C	1.600	-4.530	7.730	1.000
	KN 2x - B	KN 2x - A	1.200	-4.930	7.330	1.000
	KN 2x - C	KN 2x - A	0.178	-6.120	6.476	1.000
	KN 2x - C	KN 2x - B	-1.022	-7.320	5.276	1.000
Flowering Time (Days)	KN 4x - B	KN 4x - A	-1.250	-7.752	5.252	1.000
	KN 4x - C	KN 4x - A	-2.000	-8.298	4.298	1.000
	KN 4x - C	KN 4x - B	-0.750	-7.411	5.911	1.000
	LER 2x - B	LER 2x - A	-0.500	-6.630	5.630	1.000
	LER 2x - C	LER 2x - A	0.589	-5.709	6.887	1.000
	LER 2x - C	LER 2x - B	1.089	-5.209	7.387	1.000
	LER 4x - B	LER 4x - A	1.016	-5.892	7.924	1.000
	LER 4x - C	LER 4x - A	1.471	-5.284	8.226	1.000
	LER 4x - C	LER 4x - B	0.456	-5.843	6.754	1.000
	MT 2x - B	MT 2x - A	0.200	-5.930	6.330	1.000
	MT 2x - C	MT 2x - A	-0.500	-6.630	5.630	1.000
	MT 2x - C	MT 2x - B	-0.700	-6.830	5.430	1.000
	MT 4x - B	MT 4x - A	0.800	-5.330	6.930	1.000
	MT 4x - C	MT 4x - A	-0.500	-6.630	5.630	1.000
	MT 4x - C	MT 4x - B	-1.300	-7.430	4.830	1.000
	NO 2x - C	NO 2x - B	0.600	-5.530	6.730	1.000
	NO 4x - B	NO 4x - A	1.400	-4.730	7.530	1.000
	NO 4x - C	NO 4x - A	-0.100	-6.230	6.030	1.000
	NO 4x - D	NO 4x - A	0.300	-5.830	6.430	1.000

	NO 4x - C	NO 4x - B	-1.500	-7.630	4.630	1.000
	NO 4x - D	NO 4x - B	-1.100	-7.230	5.030	1.000
	NO 4x - D	NO 4x - C	0.400	-5.730	6.530	1.000
	WU 2x - B	WU 2x - A	-0.500	-6.630	5.630	1.000
	WU 2x - C	WU 2x - A	-1.700	-7.830	4.430	1.000
	WU 2x - C	WU 2x - B	-1.200	-7.330	4.930	1.000
	WU 4x - C	WU 4x - A	-1.000	-7.130	5.130	1.000

Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Rosette Diameter (cm)	COL 2x - B	COL 2x - A	0.160	-1.310	1.630	1.000
	COL 2x - C	COL 2x - A	0.260	-1.210	1.730	1.000
	COL 2x - C	COL 2x - B	0.100	-1.370	1.570	1.000
	COL 4x - B	COL 4x - A	0.180	-1.290	1.650	1.000
	COL 4x - C	COL 4x - A	0.070	-1.400	1.540	1.000
	COL 4x - C	COL 4x - B	-0.110	-1.580	1.360	1.000
	CT 2x - C	CT 2x - B	-0.987	-2.497	0.524	1.000
	CT 4x - B	CT 4x - A	0.040	-1.430	1.510	1.000
	CT 4x - C	CT 4x - A	0.120	-1.350	1.590	1.000
	CT 4x - D	CT 4x - A	0.070	-1.400	1.540	1.000
	CT 4x - C	CT 4x - B	0.080	-1.390	1.550	1.000
	CT 4x - D	CT 4x - B	0.030	-1.440	1.500	1.000
	CT 4x - D	CT 4x - C	-0.050	-1.520	1.420	1.000
	KN 2x - B	KN 2x - A	-0.110	-1.580	1.360	1.000
	KN 2x - C	KN 2x - A	-0.234	-1.745	1.276	1.000
	KN 2x - C	KN 2x - B	-0.124	-1.635	1.386	1.000
	KN 4x - B	KN 4x - A	-0.020	-1.490	1.450	1.000
	KN 4x - C	KN 4x - A	0.260	-1.210	1.730	1.000
	KN 4x - C	KN 4x - B	0.280	-1.190	1.750	1.000
	LER 2x - B	LER 2x - A	0.080	-1.390	1.550	1.000
	LER 2x - C	LER 2x - A	-0.156	-1.666	1.355	1.000
	LER 2x - C	LER 2x - B	-0.236	-1.746	1.275	1.000
	LER 4x - B	LER 4x - A	0.108	-1.549	1.765	1.000
	LER 4x - C	LER 4x - A	0.136	-1.484	1.756	1.000
	LER 4x - C	LER 4x - B	0.028	-1.483	1.538	1.000
	MT 2x - B	MT 2x - A	-0.210	-1.680	1.260	1.000
	MT 2x - C	MT 2x - A	-0.138	-1.648	1.373	1.000
	MT 2x - C	MT 2x - B	0.072	-1.438	1.583	1.000
	MT 4x - B	MT 4x - A	-0.321	-1.832	1.189	1.000
	MT 4x - C	MT 4x - A	-0.378	-1.928	1.172	1.000
	MT 4x - C	MT 4x - B	-0.057	-1.567	1.454	1.000
	NO 2x - C	NO 2x - B	0.322	-1.228	1.872	1.000
	NO 4x - B	NO 4x - A	-0.540	-2.010	0.930	1.000
	NO 4x - C	NO 4x - A	0.010	-1.460	1.480	1.000
	NO 4x - D	NO 4x - A	-0.000	-1.470	1.470	1.000
	NO 4x - C	NO 4x - B	0.550	-0.920	2.020	1.000

	NO 4x - D	NO 4x - B	0.540	-0.930	2.010	1.000
	NO 4x - D	NO 4x - C	-0.010	-1.480	1.460	1.000
	WU 2x - B	WU 2x - A	0.030	-1.440	1.500	1.000
	WU 2x - C	WU 2x - A	0.350	-1.120	1.820	1.000
	WU 2x - C	WU 2x - B	0.320	-1.150	1.790	1.000
	WU 4x - C	WU 4x - A	0.190	-1.280	1.660	1.000

Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Number of Leaves (15 days post germination)	COL 2x - B	COL 2x - A	-0.100	-2.893	2.693	1.000
	COL 2x - C	COL 2x - A	0.100	-2.693	2.893	1.000
	COL 2x - C	COL 2x - B	0.200	-2.593	2.993	1.000
	COL 4x - B	COL 4x - A	-0.100	-2.893	2.693	1.000
	COL 4x - C	COL 4x - A	0.100	-2.693	2.893	1.000
	COL 4x - C	COL 4x - B	0.200	-2.593	2.993	1.000
	CT 2x - C	CT 2x - B	-1.078	-3.947	1.792	1.000
	CT 4x - B	CT 4x - A	-0.200	-2.993	2.593	1.000
	CT 4x - C	CT 4x - A	0.100	-2.693	2.893	1.000
	CT 4x - D	CT 4x - A	-0.100	-2.893	2.693	1.000
	CT 4x - C	CT 4x - B	0.300	-2.493	3.093	1.000
	CT 4x - D	CT 4x - B	0.100	-2.693	2.893	1.000
	CT 4x - D	CT 4x - C	-0.200	-2.993	2.593	1.000
	KN 2x - B	KN 2x - A	0.500	-2.293	3.293	1.000
	KN 2x - C	KN 2x - A	-0.633	-3.503	2.236	1.000
	KN 2x - C	KN 2x - B	-1.133	-4.003	1.736	1.000
	KN 4x - B	KN 4x - A	-0.300	-3.093	2.493	1.000
	KN 4x - C	KN 4x - A	0.600	-2.193	3.393	1.000
	KN 4x - C	KN 4x - B	0.900	-1.893	3.693	1.000
	LER 2x - B	LER 2x - A	0.500	-2.293	3.293	1.000
	LER 2x - C	LER 2x - A	-0.322	-3.192	2.547	1.000
	LER 2x - C	LER 2x - B	-0.822	-3.692	2.047	1.000
	LER 4x - B	LER 4x - A	-0.317	-3.465	2.830	1.000
	LER 4x - C	LER 4x - A	0.371	-2.706	3.449	1.000
	LER 4x - C	LER 4x - B	0.689	-2.181	3.559	1.000
	MT 2x - B	MT 2x - A	-0.600	-3.393	2.193	1.000
	MT 2x - C	MT 2x - A	0.256	-2.614	3.125	1.000
	MT 2x - C	MT 2x - B	0.856	-2.014	3.725	1.000
	MT 4x - B	MT 4x - A	-0.344	-3.214	2.525	1.000
	MT 4x - C	MT 4x - A	-1.000	-3.944	1.944	1.000
	MT 4x - C	MT 4x - B	-0.656	-3.525	2.214	1.000
	NO 2x - C	NO 2x - B	0.667	-2.278	3.611	1.000
	NO 4x - B	NO 4x - A	-1.000	-3.793	1.793	1.000
	NO 4x - C	NO 4x - A	-0.300	-3.093	2.493	1.000
	NO 4x - D	NO 4x - A	-0.300	-3.093	2.493	1.000
	NO 4x - C	NO 4x - B	0.700	-2.093	3.493	1.000
	NO 4x - D	NO 4x - B	0.700	-2.093	3.493	1.000

	NO 4x - D	NO 4x - C	-0.000	-2.793	2.793	1.000
	WU 2x - B	WU 2x - A	0.800	-1.993	3.593	1.000
	WU 2x - C	WU 2x - A	0.700	-2.093	3.493	1.000
	WU 2x - C	WU 2x - B	-0.100	-2.893	2.693	1.000
	WU 4x - C	WU 4x - A	-0.200	-2.993	2.593	1.000

Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
	COL 2x - B	COL 2x - A	-0.468	-2.064	1.128	1.000
	COL 2x - C	COL 2x - A	-0.667	-2.263	0.929	1.000
	COL 2x - C	COL 2x - B	-0.198	-1.794	1.397	1.000
	COL 4x - B	COL 4x - A	0.266	-1.330	1.862	1.000
	COL 4x - C	COL 4x - A	0.060	-1.580	1.699	1.000
	COL 4x - C	COL 4x - B	-0.207	-1.846	1.433	1.000
	CT 2x - C	CT 2x - B	-0.026	-1.666	1.613	1.000
	CT 4x - B	CT 4x - A	0.513	-1.083	2.109	1.000
	CT 4x - C	CT 4x - A	0.702	-0.894	2.298	1.000
	CT 4x - D	CT 4x - A	-0.203	-1.962	1.555	1.000
	CT 4x - C	CT 4x - B	0.189	-1.407	1.785	1.000
	CT 4x - D	CT 4x - B	-0.716	-2.475	1.042	1.000
	CT 4x - D	CT 4x - C	-0.905	-2.664	0.853	1.000
	KN 2x - B	KN 2x - A	0.132	-1.560	1.825	1.000
	KN 2x - C	KN 2x - A	0.284	-1.312	1.880	1.000
	KN 2x - C	KN 2x - B	0.151	-1.541	1.844	1.000
	KN 4x - B	KN 4x - A	-0.190	-1.883	1.502	1.000
	KN 4x - C	KN 4x - A	0.493	-1.435	2.420	1.000
Leaf Area (cm2)	KN 4x - C	KN 4x - B	0.683	-1.160	2.526	1.000
	LER 2x - B	LER 2x - A	0.228	-1.412	1.867	1.000
	LER 2x - C	LER 2x - A	0.816	-0.824	2.455	1.000
	LER 2x - C	LER 2x - B	0.588	-1.094	2.270	1.000
	LER 4x - B	LER 4x - A	-0.280	-2.207	1.647	1.000
	LER 4x - C	LER 4x - A	-0.050	-1.893	1.792	1.000
	LER 4x - C	LER 4x - B	0.230	-1.463	1.922	1.000
	MT 2x - B	MT 2x - A	-0.278	-1.874	1.318	1.000
	MT 2x - C	MT 2x - A	-0.316	-1.912	1.280	1.000
	MT 2x - C	MT 2x - B	-0.038	-1.634	1.558	1.000
	MT 4x - B	MT 4x - A	-0.142	-1.825	1.540	1.000
	MT 4x - C	MT 4x - A	-0.324	-1.964	1.315	1.000
	MT 4x - C	MT 4x - B	-0.182	-1.822	1.458	1.000
	NO 2x - C	NO 2x - B	-0.061	-1.657	1.535	1.000
	NO 4x - B	NO 4x - A	-0.479	-2.171	1.214	1.000
	NO 4x - C	NO 4x - A	-0.405	-2.044	1.235	1.000
	NO 4x - D	NO 4x - A	-1.123	-2.816	0.569	0.999
	NO 4x - C	NO 4x - B	0.074	-1.660	1.808	1.000
	NO 4x - D	NO 4x - B	-0.645	-2.429	1.140	1.000
	NO 4x - D	NO 4x - C	-0.718	-2.452	1.016	1.000

	WU 2x - B	WU 2x - A	-0.193	-1.788	1.403	1.000
	WU 2x - C	WU 2x - A	0.507	-1.089	2.102	1.000
	WU 2x - C	WU 2x - B	0.699	-0.897	2.295	1.000
	WU 4x - C	WU 4x - A	0.730	-0.909	2.370	1.000

Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Trichome Density (trichomes per cm2)	COL 2x - B	COL 2x - A	6.443	-30.343	43.228	1.000
	COL 2x - C	COL 2x - A	19.003	-17.782	55.789	1.000
	COL 2x - C	COL 2x - B	12.561	-23.243	48.365	1.000
	COL 4x - B	COL 4x - A	4.085	-31.719	39.889	1.000
	COL 4x - C	COL 4x - A	6.397	-30.389	43.182	1.000
	COL 4x - C	COL 4x - B	2.311	-34.474	39.097	1.000
	CT 2x - C	CT 2x - B	3.589	-33.197	40.374	1.000
	CT 4x - B	CT 4x - A	-9.057	-47.959	29.846	1.000
	CT 4x - C	CT 4x - A	-13.641	-51.382	24.100	1.000
	CT 4x - D	CT 4x - A	11.432	-28.915	51.779	1.000
	CT 4x - C	CT 4x - B	-4.585	-43.487	34.318	1.000
	CT 4x - D	CT 4x - B	20.489	-20.947	61.924	1.000
	CT 4x - D	CT 4x - C	25.073	-15.273	65.420	1.000
	KN 2x - B	KN 2x - A	-4.143	-44.173	35.887	1.000
	KN 2x - C	KN 2x - A	-7.934	-45.910	30.043	1.000
	KN 2x - C	KN 2x - B	-3.790	-41.767	34.186	1.000
	KN 4x - B	KN 4x - A	7.226	-30.750	45.202	1.000
	KN 4x - C	KN 4x - A	-5.161	-48.399	38.076	1.000
	KN 4x - C	KN 4x - B	-12.388	-53.731	28.956	1.000
	LER 2x - B	LER 2x - A	0.561	-36.224	37.346	1.000
	LER 2x - C	LER 2x - A	-2.098	-38.884	34.687	1.000
	LER 2x - C	LER 2x - B	-2.659	-40.400	35.082	1.000
	LER 4x - B	LER 4x - A	8.240	-34.997	51.478	1.000
	LER 4x - C	LER 4x - A	-0.338	-41.681	41.005	1.000
	LER 4x - C	LER 4x - B	-8.578	-46.554	29.398	1.000
	MT 2x - B	MT 2x - A	7.581	-28.224	43.385	1.000
	MT 2x - C	MT 2x - A	13.746	-22.058	49.550	1.000
	MT 2x - C	MT 2x - B	6.166	-29.639	41.970	1.000
	MT 4x - B	MT 4x - A	3.466	-34.275	41.207	1.000
	MT 4x - C	MT 4x - A	20.419	-16.367	57.204	1.000
	MT 4x - C	MT 4x - B	16.952	-19.833	53.737	1.000
	NO 2x - C	NO 2x - B	2.425	-33.380	38.229	1.000
	NO 4x - B	NO 4x - A	-2.971	-40.947	35.005	1.000
	NO 4x - C	NO 4x - A	0.594	-36.191	37.380	1.000
	NO 4x - D	NO 4x - A	9.462	-28.514	47.438	1.000
	NO 4x - C	NO 4x - B	3.565	-35.337	42.468	1.000
	NO 4x - D	NO 4x - B	12.433	-27.597	52.464	1.000
	NO 4x - D	NO 4x - C	8.868	-30.035	47.770	1.000
	WU 2x - B	WU 2x - A	4.550	-31.254	40.354	1.000

	WU 2x - C	WU 2x - A	-4.724	-40.528	31.080	1.000
	WU 2x - C	WU 2x - B	-9.274	-45.078	26.530	1.000
	WU 4x - C	WU 4x - A	-1.735	-38.521	35.050	1.000

Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Total Trichomes (per leaf)	COL 2x - B	COL 2x - A	2.044	-88.089	92.178	1.000
	COL 2x - C	COL 2x - A	14.944	-75.189	105.078	1.000
	COL 2x - C	COL 2x - B	12.900	-74.829	100.629	1.000
	COL 4x - B	COL 4x - A	24.100	-63.629	111.829	1.000
	COL 4x - C	COL 4x - A	15.611	-74.522	105.744	1.000
	COL 4x - C	COL 4x - B	-8.489	-98.622	81.644	1.000
	CT 2x - C	CT 2x - B	4.600	-85.533	94.733	1.000
	CT 4x - B	CT 4x - A	-12.458	-107.779	82.862	1.000
	CT 4x - C	CT 4x - A	-9.000	-101.475	83.475	1.000
	CT 4x - D	CT 4x - A	15.524	-83.336	114.383	1.000
	CT 4x - C	CT 4x - B	3.458	-91.862	98.779	1.000
	CT 4x - D	CT 4x - B	27.982	-73.545	129.509	1.000
	CT 4x - D	CT 4x - C	24.524	-74.336	123.383	1.000
	KN 2x - B	KN 2x - A	-2.500	-100.584	95.584	1.000
	KN 2x - C	KN 2x - A	-5.375	-98.426	87.676	1.000
	KN 2x - C	KN 2x - B	-2.875	-95.926	90.176	1.000
	KN 4x - B	KN 4x - A	3.475	-89.576	96.526	1.000
	KN 4x - C	KN 4x - A	-9.458	-115.401	96.485	1.000
	KN 4x - C	KN 4x - B	-12.933	-114.234	88.368	1.000
	LER 2x - B	LER 2x - A	6.433	-83.700	96.566	1.000
	LER 2x - C	LER 2x - A	21.767	-68.366	111.900	1.000
	LER 2x - C	LER 2x - B	15.333	-77.141	107.808	1.000
	LER 4x - B	LER 4x - A	9.875	-96.068	115.818	1.000
	LER 4x - C	LER 4x - A	-3.100	-104.401	98.201	1.000
	LER 4x - C	LER 4x - B	-12.975	-106.026	80.076	1.000
	MT 2x - B	MT 2x - A	-10.700	-98.429	77.029	1.000
	MT 2x - C	MT 2x - A	1.900	-85.829	89.629	1.000
	MT 2x - C	MT 2x - B	12.600	-75.129	100.329	1.000
	MT 4x - B	MT 4x - A	-0.222	-92.697	92.252	1.000
	MT 4x - C	MT 4x - A	18.033	-72.100	108.166	1.000
	MT 4x - C	MT 4x - B	18.256	-71.878	108.389	1.000
	NO 2x - C	NO 2x - B	12.700	-75.029	100.429	1.000
	NO 4x - B	NO 4x - A	-21.875	-114.926	71.176	1.000
	NO 4x - C	NO 4x - A	-8.444	-98.578	81.689	1.000
	NO 4x - D	NO 4x - A	-5.250	-98.301	87.801	1.000
	NO 4x - C	NO 4x - B	13.431	-81.890	108.751	1.000
	NO 4x - D	NO 4x - B	16.625	-81.459	114.709	1.000
	NO 4x - D	NO 4x - C	3.194	-92.126	98.515	1.000
	WU 2x - B	WU 2x - A	-5.300	-93.029	82.429	1.000
	WU 2x - C	WU 2x - A	19.400	-68.329	107.129	1.000

	WU 2x - C	WU 2x - B	24.700	-63.029	112.429	1.000
	WU 4x - C	WU 4x - A	60.644	-29.489	150.778	0.998

Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Trichomes with >3 branches (% of total trichomes on the leaf)	COL 2x - B	COL 2x - A	1.032	-24.904	26.969	1.000
	COL 2x - C	COL 2x - A	-0.062	-25.998	25.875	1.000
	COL 2x - C	COL 2x - B	-1.094	-26.338	24.151	1.000
	COL 4x - B	**COL 4x - A**	**-26.466**	**-51.710**	**-1.221**	**0.018**
	COL 4x - C	COL 4x - A	-2.655	-28.592	23.281	1.000
	COL 4x - C	COL 4x - B	23.810	-2.126	49.747	0.215
	CT 2x - C	CT 2x - B	7.100	-18.837	33.036	1.000
	CT 4x - B	CT 4x - A	0.967	-26.462	28.396	1.000
	CT 4x - C	CT 4x - A	8.119	-18.491	34.729	1.000
	CT 4x - D	CT 4x - A	10.777	-17.671	39.224	1.000
	CT 4x - C	CT 4x - B	7.152	-20.277	34.581	1.000
	CT 4x - D	CT 4x - B	9.810	-19.405	39.025	1.000
	CT 4x - D	CT 4x - C	2.658	-25.790	31.105	1.000
	KN 2x - B	KN 2x - A	-3.479	-31.703	24.746	1.000
	KN 2x - C	KN 2x - A	-0.657	-27.433	26.119	1.000
	KN 2x - C	KN 2x - B	2.822	-23.954	29.598	1.000
	KN 4x - B	KN 4x - A	0.619	-26.157	27.394	1.000
	KN 4x - C	KN 4x - A	1.648	-28.838	32.133	1.000
	KN 4x - C	KN 4x - B	1.029	-28.121	30.179	1.000
	LER 2x - B	LER 2x - A	-5.823	-31.759	20.114	1.000
	LER 2x - C	LER 2x - A	2.126	-23.810	28.062	1.000
	LER 2x - C	LER 2x - B	7.949	-18.661	34.559	1.000
	LER 4x - B	LER 4x - A	-19.699	-50.184	10.787	1.000
	LER 4x - C	LER 4x - A	1.241	-27.909	30.391	1.000
	LER 4x - C	LER 4x - B	20.940	-5.836	47.716	0.837
	MT 2x - B	MT 2x - A	0.456	-24.788	25.701	1.000
	MT 2x - C	MT 2x - A	0.020	-25.224	25.265	1.000
	MT 2x - C	MT 2x - B	-0.436	-25.681	24.809	1.000
	MT 4x - B	MT 4x - A	-4.181	-30.791	22.429	1.000
	MT 4x - C	MT 4x - A	2.567	-23.369	28.504	1.000
	MT 4x - C	MT 4x - B	6.748	-19.188	32.684	1.000
	NO 2x - C	NO 2x - B	0.142	-25.103	25.386	1.000
	NO 4x - B	NO 4x - A	0.110	-26.666	26.886	1.000
	NO 4x - C	NO 4x - A	2.774	-23.162	28.710	1.000
	NO 4x - D	NO 4x - A	-2.811	-29.586	23.965	1.000
	NO 4x - C	NO 4x - B	2.664	-24.765	30.093	1.000
	NO 4x - D	NO 4x - B	-2.921	-31.145	25.303	1.000
	NO 4x - D	NO 4x - C	-5.585	-33.014	21.845	1.000
	WU 2x - B	WU 2x - A	-5.443	-30.688	19.801	1.000
	WU 2x - C	WU 2x - A	3.522	-21.722	28.767	1.000
	WU 2x - C	WU 2x - B	8.966	-16.279	34.210	1.000

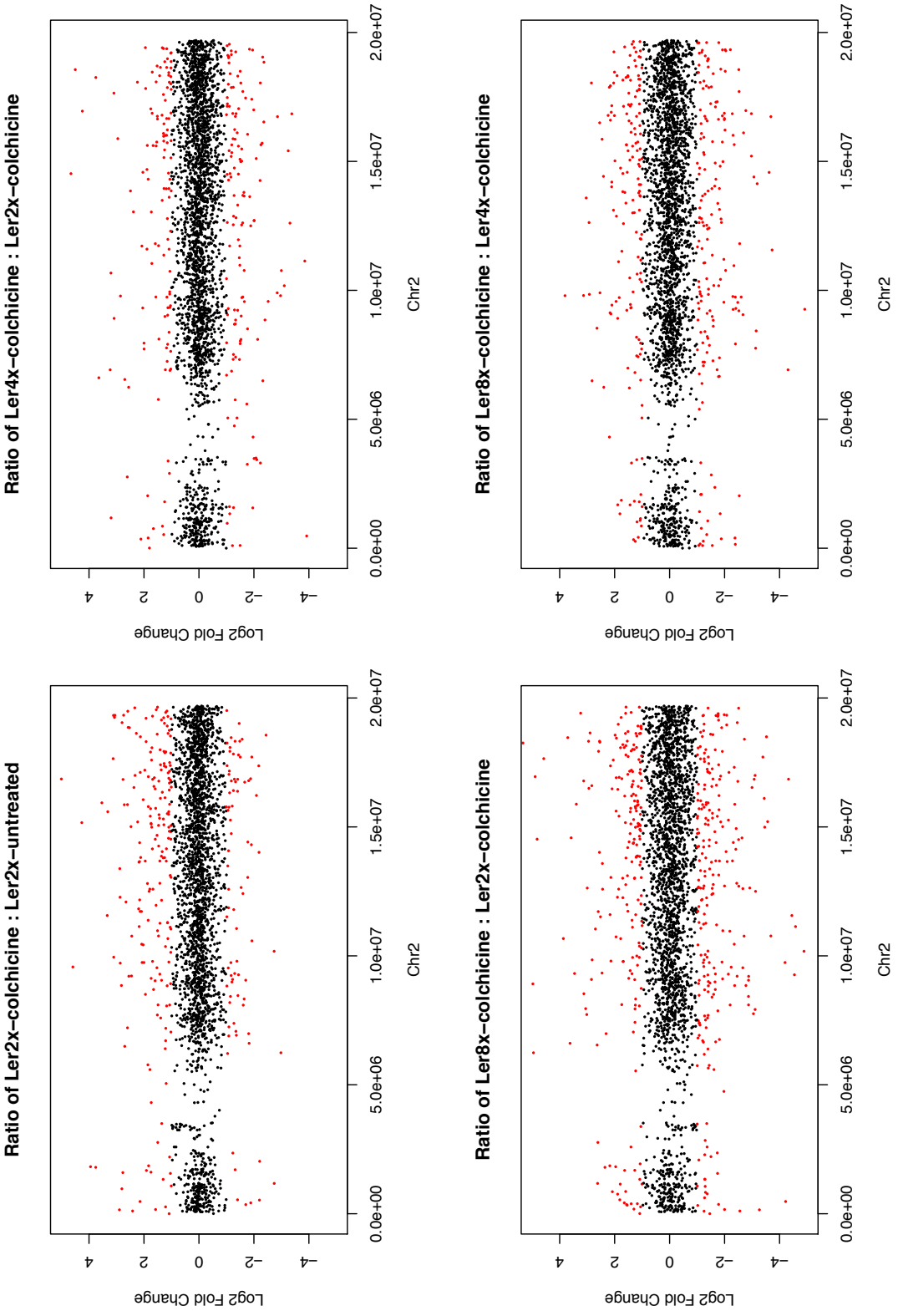
	WU 4x - C	WU 4x - A	5.270	-20.666	31.207	1.000

Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Stomatal Density (per mm2)	COL 2x - B	COL 2x - A	0.471	-61.061	62.003	1.000
	COL 2x - C	COL 2x - A	24.956	-36.576	86.489	1.000
	COL 2x - C	COL 2x - B	24.486	-37.047	86.018	1.000
	COL 4x - B	COL 4x - A	4.761	-58.457	67.979	1.000
	COL 4x - C	COL 4x - A	8.005	-53.527	69.537	1.000
	COL 4x - C	COL 4x - B	3.244	-59.974	66.462	1.000
	CT 2x - C	CT 2x - B	54.739	-29.517	138.996	1.000
	CT 4x - B	CT 4x - A	-5.651	-67.183	55.882	1.000
	CT 4x - C	CT 4x - A	-6.121	-67.654	55.411	1.000
	CT 4x - D	CT 4x - A	-7.534	-69.066	53.998	1.000
	CT 4x - C	CT 4x - B	-0.471	-62.003	61.061	1.000
	CT 4x - D	CT 4x - B	-1.884	-63.416	59.649	1.000
	CT 4x - D	CT 4x - C	-1.413	-62.945	60.120	1.000
	KN 2x - B	KN 2x - A	-29.037	-95.894	37.819	1.000
	KN 2x - C	KN 2x - A	-21.503	-84.722	41.715	1.000
	KN 2x - C	KN 2x - B	7.534	-57.731	72.799	1.000
	KN 4x - B	KN 4x - A	8.005	-53.527	69.537	1.000
	KN 4x - C	KN 4x - A	1.884	-59.649	63.416	1.000
	KN 4x - C	KN 4x - B	-6.121	-67.654	55.411	1.000
	LER 2x - B	LER 2x - A	13.185	-50.034	76.403	1.000
	LER 2x - C	LER 2x - A	52.947	-10.271	116.166	0.578
	LER 2x - C	LER 2x - B	39.763	-25.098	104.623	1.000
	LER 4x - B	LER 4x - A	-61.803	-207.739	84.134	1.000
	LER 4x - C	LER 4x - A	-48.971	-193.277	95.335	1.000
	LER 4x - C	LER 4x - B	12.831	-52.433	78.096	1.000
	MT 2x - B	MT 2x - A	11.458	-59.593	82.509	1.000
	MT 2x - C	MT 2x - A	31.078	-30.454	92.610	1.000
	MT 2x - C	MT 2x - B	19.620	-51.431	90.671	1.000
	MT 4x - B	MT 4x - A	-4.970	-68.189	58.248	1.000
	MT 4x - C	MT 4x - A	0.000	-64.861	64.861	1.000
	MT 4x - C	MT 4x - B	4.970	-58.248	68.189	1.000
	NO 2x - C	NO 2x - B	55.799	-9.466	121.064	0.488
	NO 4x - B	NO 4x - A	10.359	-51.173	71.891	1.000
	NO 4x - C	NO 4x - A	12.714	-48.819	74.246	1.000
	NO 4x - D	NO 4x - A	-3.296	-64.828	58.236	1.000
	NO 4x - C	NO 4x - B	2.354	-59.178	63.887	1.000
	NO 4x - D	NO 4x - B	-13.655	-75.188	47.877	1.000
	NO 4x - D	NO 4x - C	-16.010	-77.542	45.522	1.000
	WU 2x - B	WU 2x - A	-0.471	-62.003	61.061	1.000
	WU 2x - C	WU 2x - A	-4.238	-65.770	57.294	1.000
	WU 2x - C	WU 2x - B	-3.767	-65.299	57.765	1.000
	WU 4x - C	WU 4x - A	8.947	-52.586	70.479	1.000

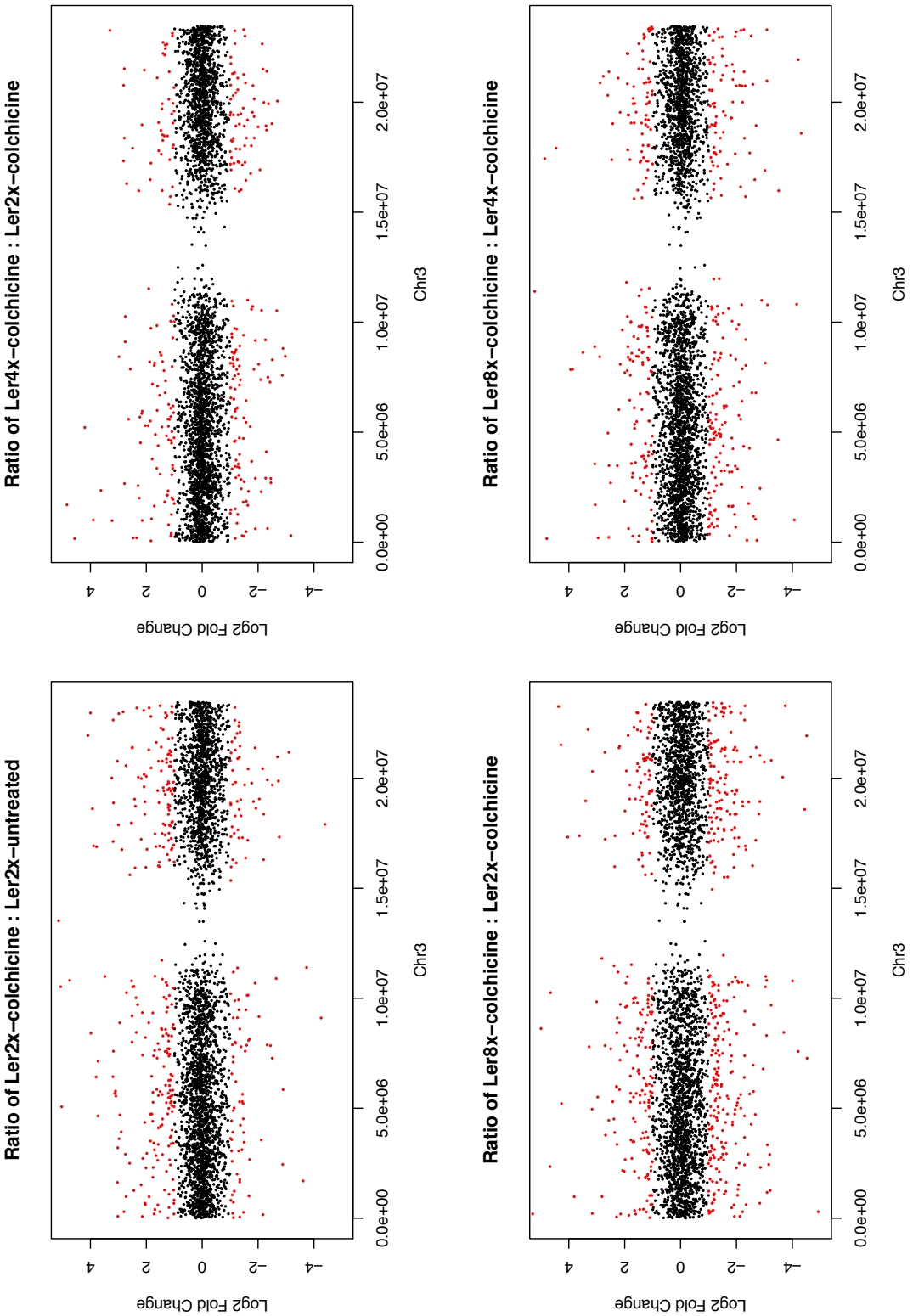
Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Stomatal Diameter (µm)	COL 2x - B	COL 2x - A	0.460	-5.915	6.835	1.000
	COL 2x - C	COL 2x - A	1.320	-5.055	7.695	1.000
	COL 2x - C	COL 2x - B	0.860	-5.515	7.235	1.000
	COL 4x - B	COL 4x - A	0.650	-5.900	7.199	1.000
	COL 4x - C	COL 4x - A	1.774	-4.601	8.149	1.000
	COL 4x - C	COL 4x - B	1.124	-5.425	7.674	1.000
	CT 2x - C	CT 2x - B	-3.303	-12.032	5.427	1.000
	CT 4x - B	CT 4x - A	0.300	-6.075	6.675	1.000
	CT 4x - C	CT 4x - A	0.244	-6.131	6.619	1.000
	CT 4x - D	CT 4x - A	0.400	-5.975	6.775	1.000
	CT 4x - C	CT 4x - B	-0.056	-6.431	6.319	1.000
	CT 4x - D	CT 4x - B	0.100	-6.275	6.475	1.000
	CT 4x - D	CT 4x - C	0.156	-6.219	6.531	1.000
	KN 2x - B	KN 2x - A	-0.659	-7.586	6.267	1.000
	KN 2x - C	KN 2x - A	-0.693	-7.242	5.857	1.000
	KN 2x - C	KN 2x - B	-0.034	-6.795	6.728	1.000
	KN 4x - B	KN 4x - A	-0.448	-6.823	5.927	1.000
	KN 4x - C	KN 4x - A	1.742	-4.633	8.117	1.000
	KN 4x - C	KN 4x - B	2.190	-4.185	8.565	1.000
	LER 2x - B	LER 2x - A	2.129	-4.420	8.679	1.000
	LER 2x - C	LER 2x - A	1.025	-5.525	7.574	1.000
	LER 2x - C	LER 2x - B	-1.104	-7.824	5.615	1.000
	LER 4x - B	LER 4x - A	0.683	-14.437	15.802	1.000
	LER 4x - C	LER 4x - A	-0.428	-15.378	14.522	1.000
	LER 4x - C	LER 4x - B	-1.111	-7.872	5.651	1.000
	MT 2x - B	MT 2x - A	-2.668	-10.476	5.140	1.000
	MT 2x - C	MT 2x - A	-0.906	-7.281	5.469	1.000
	MT 2x - C	MT 2x - B	1.762	-6.046	9.570	1.000
	MT 4x - B	MT 4x - A	-2.236	-8.955	4.484	1.000
	MT 4x - C	MT 4x - A	-0.520	-7.240	6.200	1.000
	MT 4x - C	MT 4x - B	1.716	-5.004	8.435	1.000
	NO 2x - C	NO 2x - B	-0.145	-6.906	6.617	1.000
	NO 4x - B	NO 4x - A	2.202	-4.173	8.577	1.000
	NO 4x - C	NO 4x - A	-0.741	-7.291	5.808	1.000
	NO 4x - D	NO 4x - A	0.006	-6.369	6.381	1.000
	NO 4x - C	NO 4x - B	-2.943	-9.493	3.606	1.000
	NO 4x - D	NO 4x - B	-2.196	-8.571	4.179	1.000
	NO 4x - D	NO 4x - C	0.747	-5.802	7.297	1.000
	WU 2x - B	WU 2x - A	-0.366	-6.741	6.009	1.000
	WU 2x - C	WU 2x - A	1.532	-4.843	7.907	1.000
	WU 2x - C	WU 2x - B	1.898	-4.477	8.273	1.000
	WU 4x - C	WU 4x - A	1.464	-4.911	7.839	1.000

Trait	Contrast 1 (I)	Contrast 2 (J)	Mean Diff (I-J)	Lower Bound	Upper Bound	Sig
Total Fruits	COL 2x - B	COL 2x - A	-56.600	-234.178	120.978	1.000
	COL 2x - C	COL 2x - A	-71.800	-249.378	105.778	1.000
	COL 2x - C	COL 2x - B	-15.200	-192.778	162.378	1.000
	COL 4x - B	COL 4x - A	3.800	-173.778	181.378	1.000
	COL 4x - C	COL 4x - A	-18.767	-201.211	163.677	1.000
	COL 4x - C	COL 4x - B	-22.567	-205.011	159.877	1.000
	CT 2x - C	CT 2x - B	-42.111	-224.555	140.333	1.000
	CT 4x - B	CT 4x - A	118.900	-58.678	296.478	0.999
	CT 4x - C	CT 4x - A	112.500	-65.078	290.078	1.000
	CT 4x - D	CT 4x - A	5.556	-176.888	187.999	1.000
	CT 4x - C	CT 4x - B	-6.400	-183.978	171.178	1.000
	CT 4x - D	CT 4x - B	-113.344	-295.788	69.099	1.000
	CT 4x - D	CT 4x - C	-106.944	-289.388	75.499	1.000
	KN 2x - B	KN 2x - A	10.900	-166.678	188.478	1.000
	KN 2x - C	KN 2x - A	18.800	-158.778	196.378	1.000
	KN 2x - C	KN 2x - B	7.900	-169.678	185.478	1.000
	KN 4x - B	KN 4x - A	-7.900	-185.478	169.678	1.000
	KN 4x - C	KN 4x - A	-29.600	-207.178	147.978	1.000
	KN 4x - C	KN 4x - B	-21.700	-199.278	155.878	1.000
	LER 2x - B	LER 2x - A	-33.000	-210.578	144.578	1.000
	LER 2x - C	LER 2x - A	17.711	-164.733	200.155	1.000
	LER 2x - C	LER 2x - B	50.711	-131.733	233.155	1.000
	LER 4x - B	LER 4x - A	10.381	-189.727	210.488	1.000
	LER 4x - C	LER 4x - A	70.414	-125.267	266.095	1.000
	LER 4x - C	LER 4x - B	60.033	-122.411	242.477	1.000
	MT 2x - B	MT 2x - A	-4.900	-182.478	172.678	1.000
	MT 2x - C	MT 2x - A	-3.900	-181.478	173.678	1.000
	MT 2x - C	MT 2x - B	1.000	-176.578	178.578	1.000
	MT 4x - B	MT 4x - A	-77.200	-254.778	100.378	1.000
	MT 4x - C	MT 4x - A	-21.400	-198.978	156.178	1.000
	MT 4x - C	MT 4x - B	55.800	-121.778	233.378	1.000
	NO 2x - C	NO 2x - B	9.200	-168.378	186.778	1.000
	NO 4x - B	NO 4x - A	-3.300	-180.878	174.278	1.000
	NO 4x - C	NO 4x - A	-27.600	-205.178	149.978	1.000
	NO 4x - D	NO 4x - A	-53.400	-230.978	124.178	1.000
	NO 4x - C	NO 4x - B	-24.300	-201.878	153.278	1.000
	NO 4x - D	NO 4x - B	-50.100	-227.678	127.478	1.000
	NO 4x - D	NO 4x - C	-25.800	-203.378	151.778	1.000
	WU 2x - B	WU 2x - A	-6.000	-183.578	171.578	1.000
	WU 2x - C	WU 2x - A	86.600	-90.978	264.178	1.000
	WU 2x - C	WU 2x - B	92.600	-84.978	270.178	1.000
	WU 4x - C	WU 4x - A	56.000	-121.578	233.578	1.000

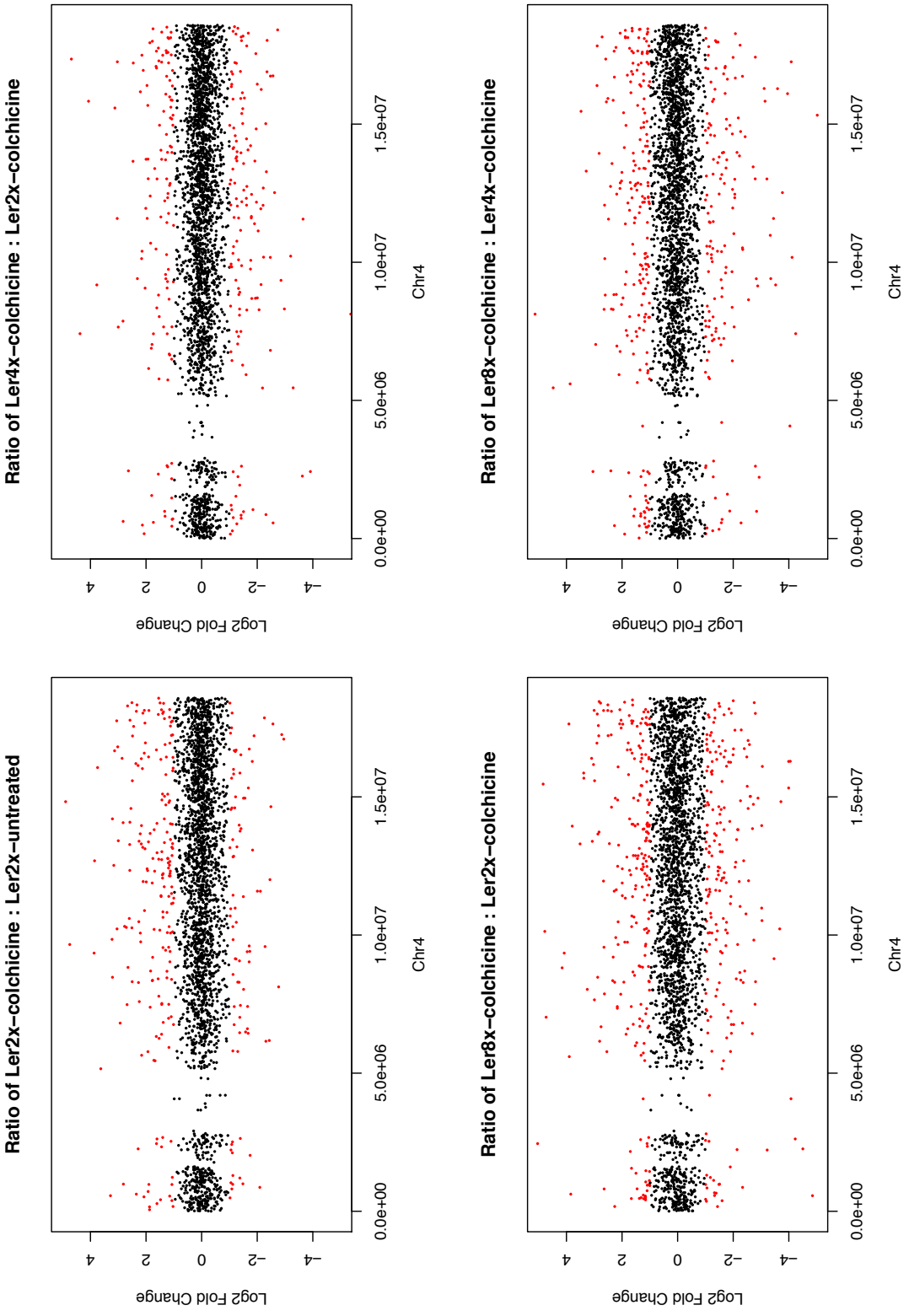
Supplementary Data 7.4 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) for Ler across chromosome2 (genomic coordinates on the x-axis). (A) 2x-colchicine to 2x-untreated (B) 4x-colchicine to 2x-colchicine (C) 8x-colchicine to 2x-colchicine (D) 8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.



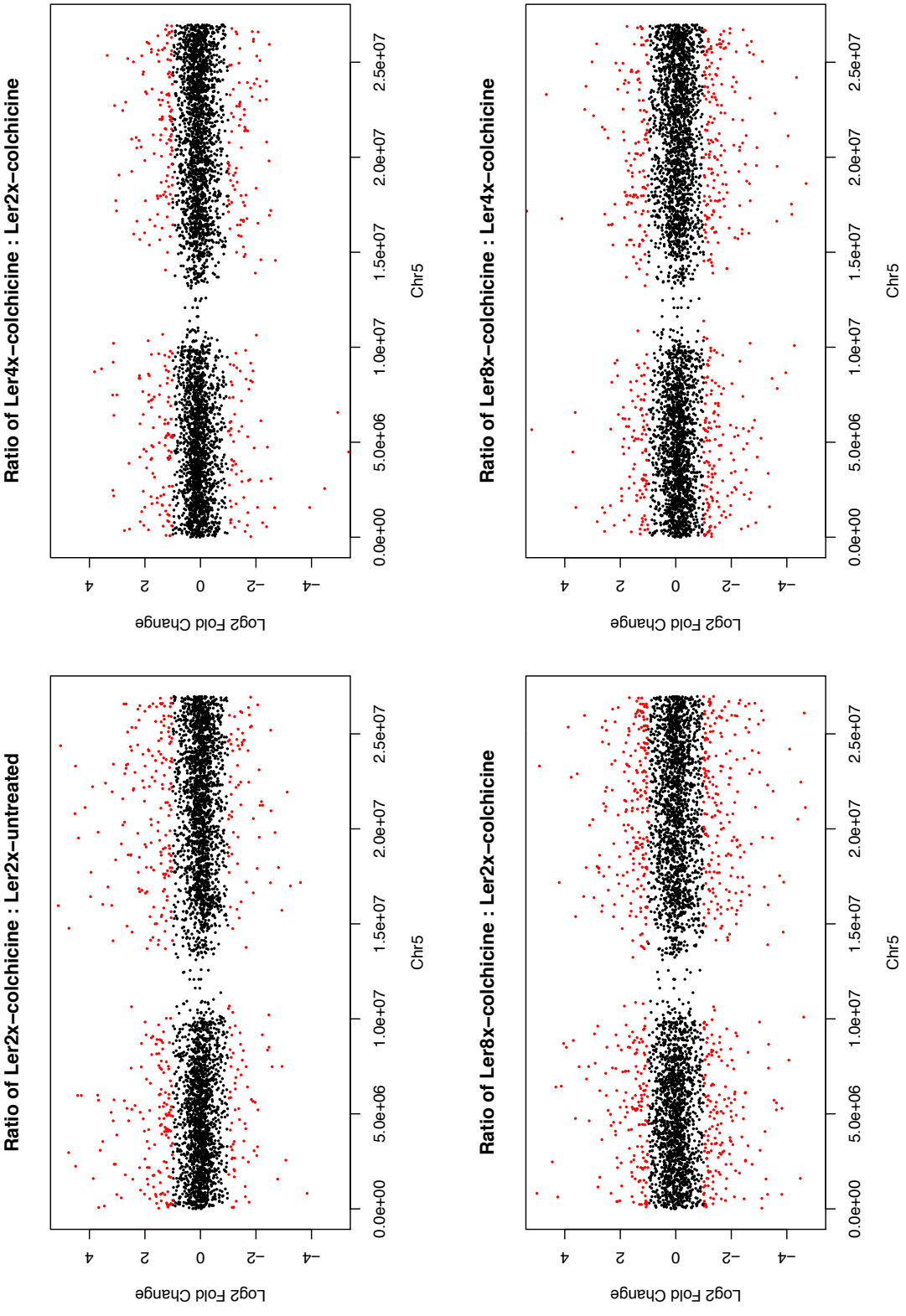
Supplementary Data 7.5 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) for Ler across chromosome3 (genomic coordinates on the x-axis). (A) 2x-colchicine to 2x-untreated (B) 4x-colchicine to 2x-colchicine (C) 8x-colchicine to 2x-colchicine (D) 8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.



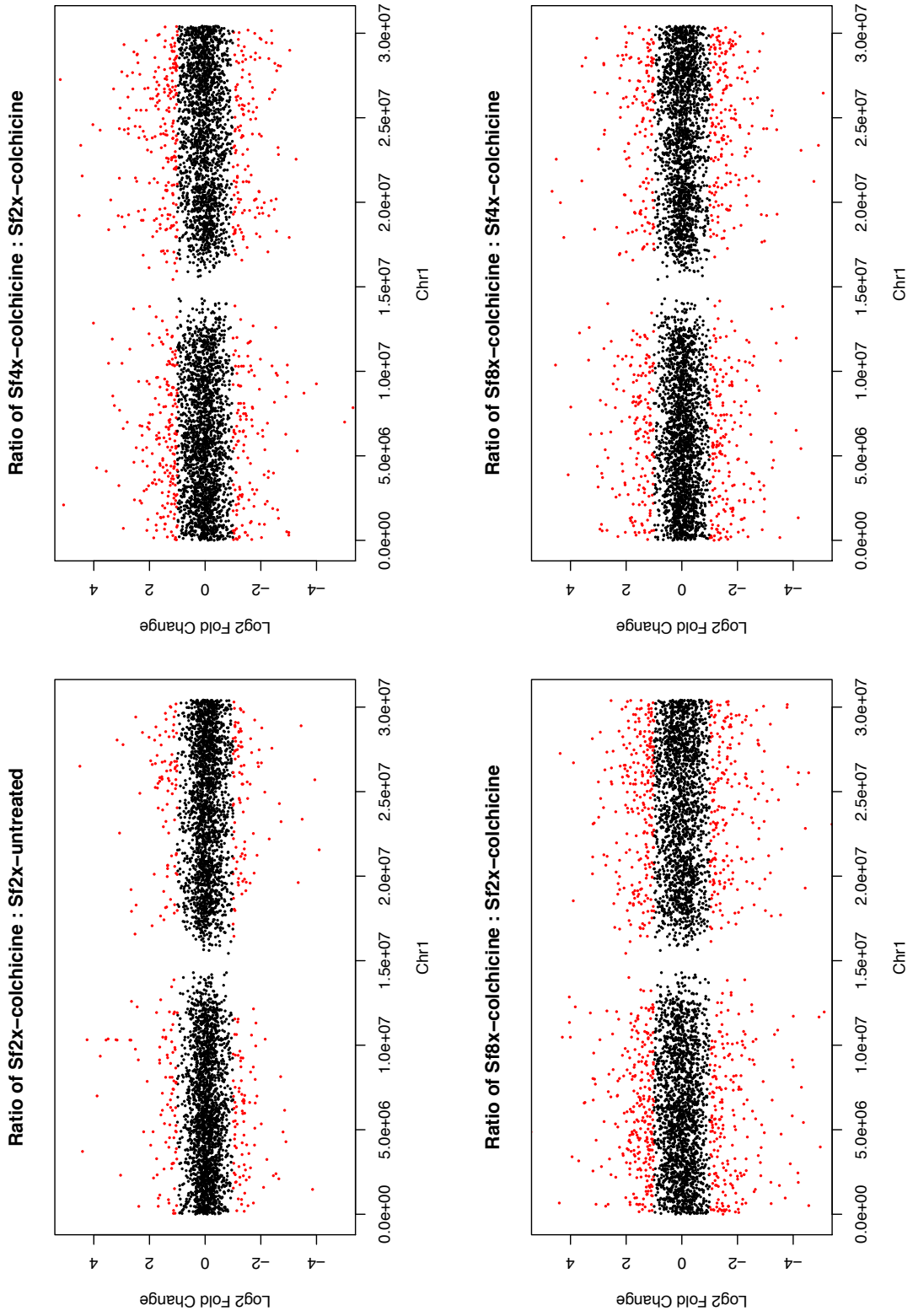
Supplementary Data 7.6 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) for Ler across chromosome4 (genomic coordinates on the x-axis). (A) 2x-colchicine to 2x-untreated (B) 4x-colchicine to 2x-colchicine (C) 8x-colchicine to 2x-colchicine (D) 8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.



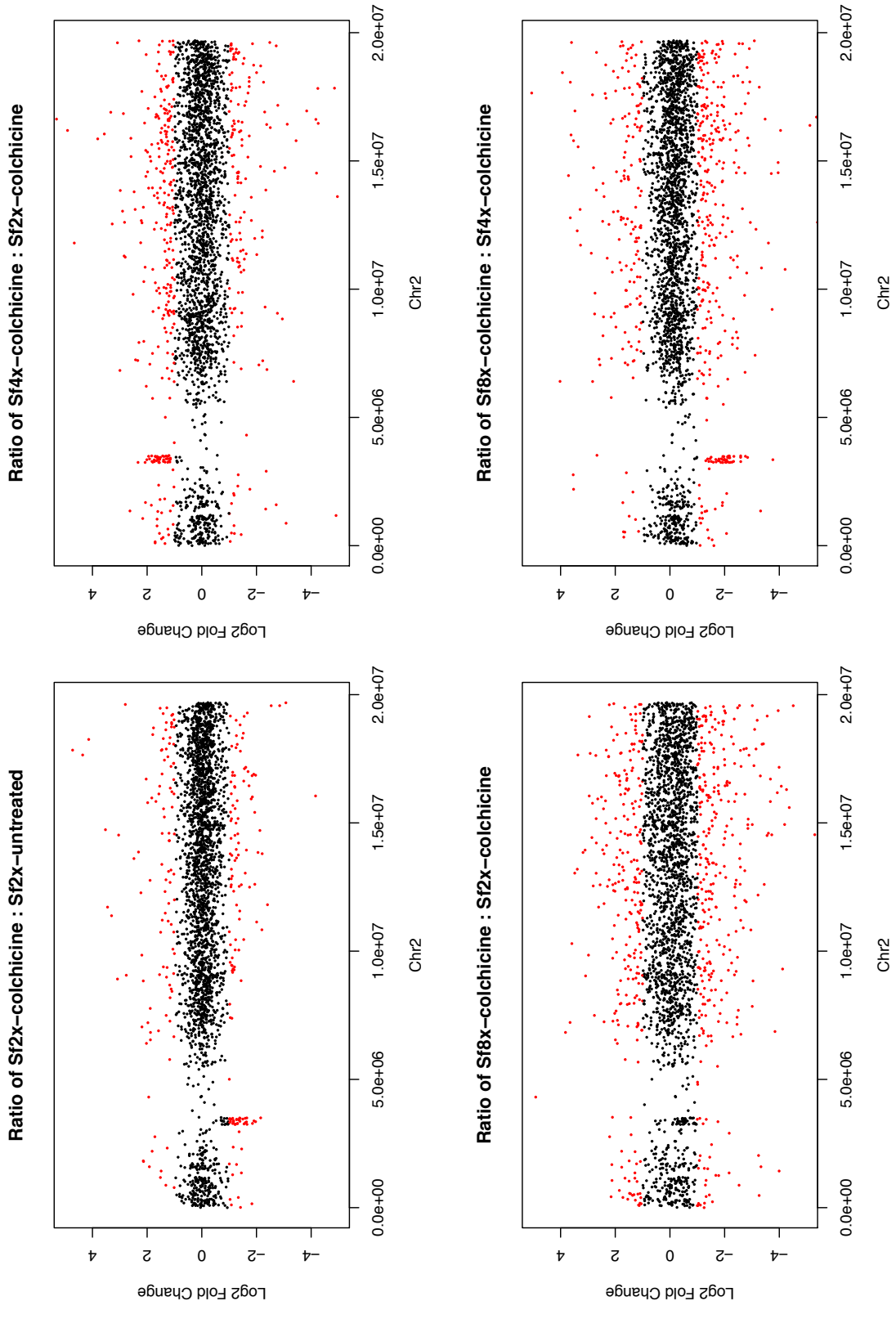
Supplementary Data 7.7 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) for Ler across chromosome5 (genomic coordinates on the x-axis). (A) 2x-colchicine to 2x-untreated (B) 4x-colchicine to 2x-colchicine (C) 8x-colchicine to 2x-colchicine (D) 8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.



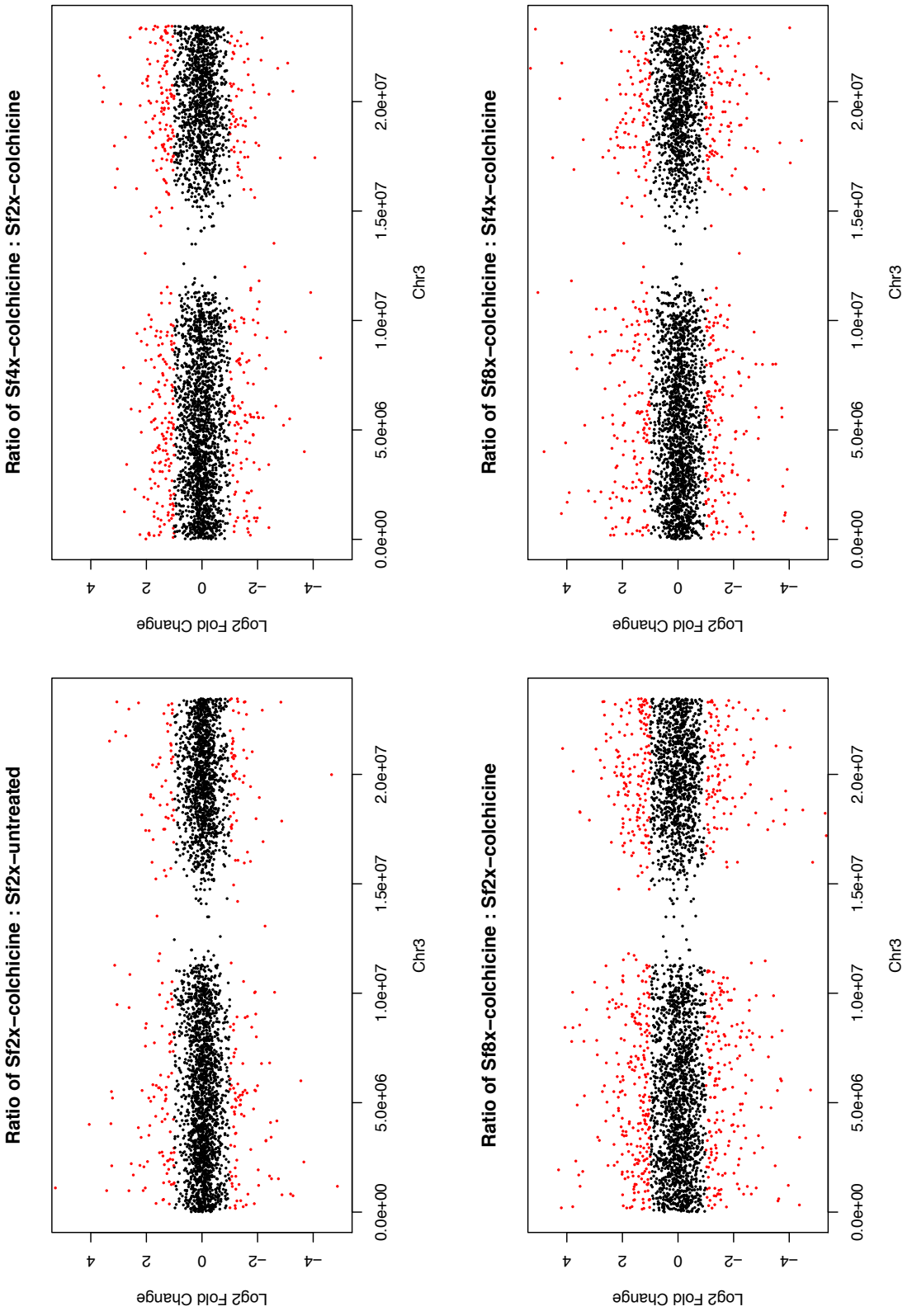
Supplementary Data 7.8 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) for Sf across chromosomes1 (genomic coordinates on the x-axis). (A) 2x-colchicine to 2x-untreated (B) 4x-colchicine to 2x-colchicine (C) 8x-colchicine to 2x-colchicine (D) 8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.



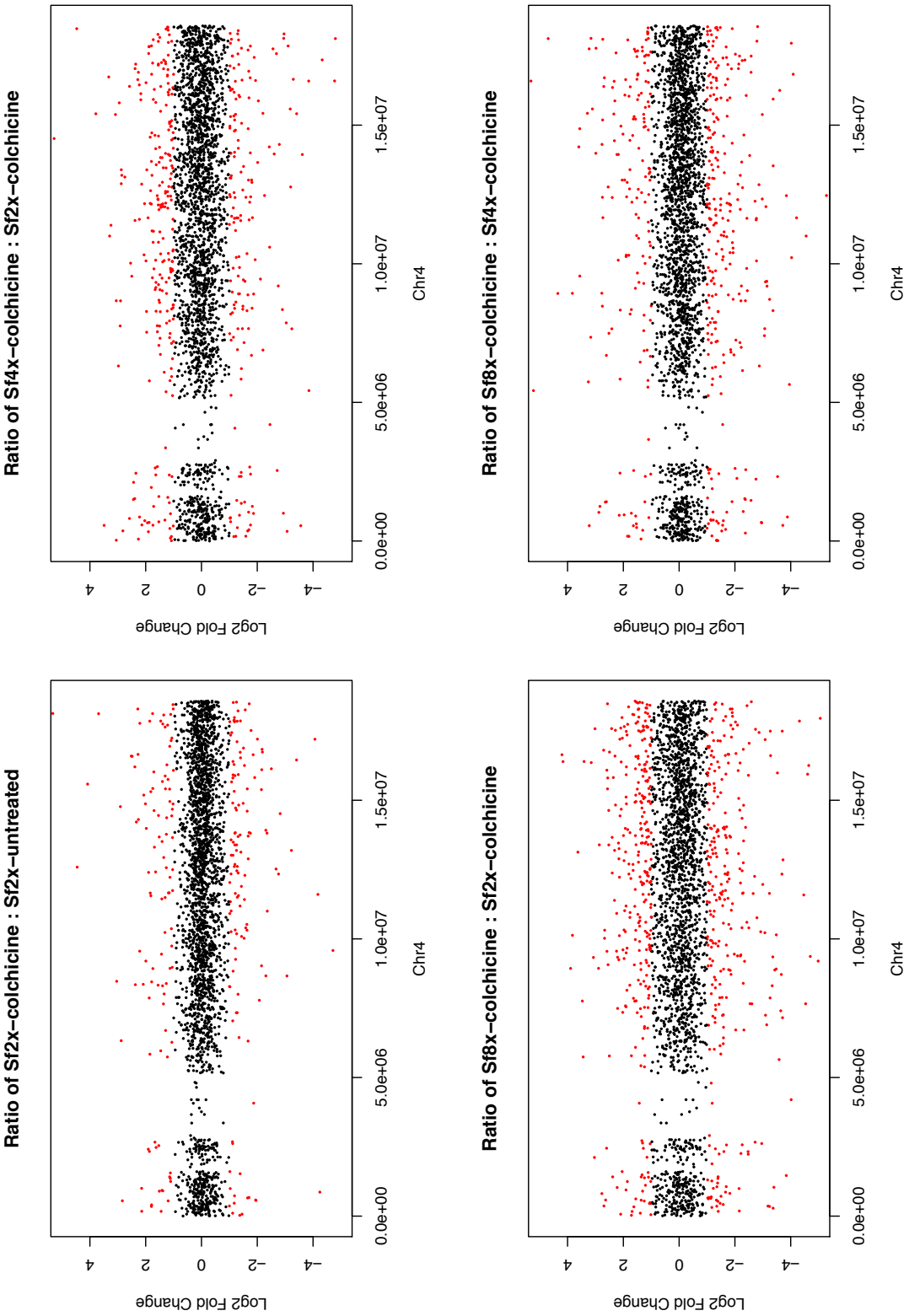
Supplementary Data 7.9 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) for Sf across chromosomes2 (genomic coordinates on the x-axis). (A) 2x-colchicine to 2x-untreated (B) 4x-colchicine to 2x-colchicine (C) 8x-colchicine to 2x-colchicine (D) 8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.



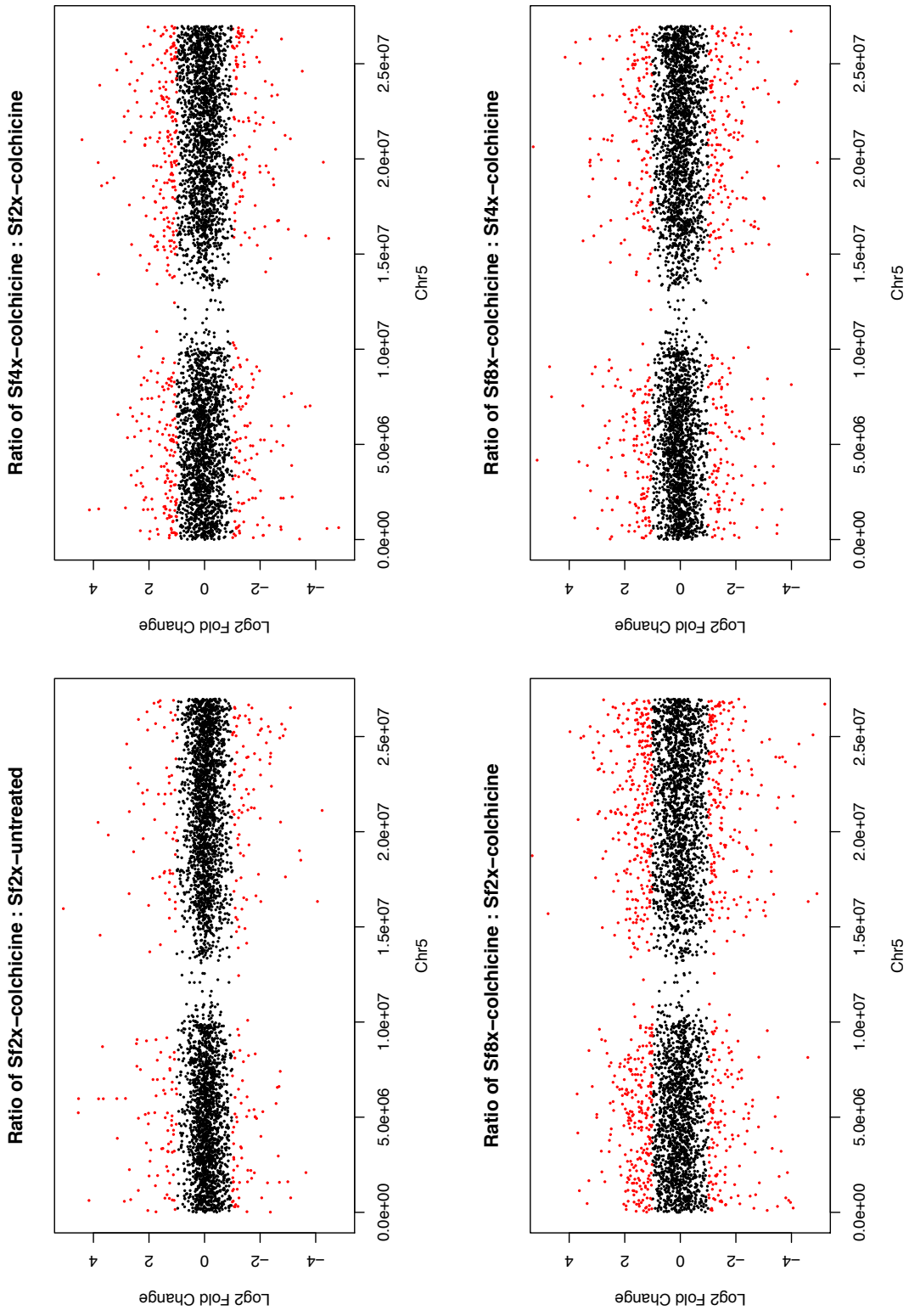
Supplementary Data 7.10 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) for Sf across chromosome3 (genomic coordinates on the x-axis). (A) 2x-colchicine to 2x-untreated (B) 4x-colchicine to 2x-colchicine (C) 8x-colchicine to 2x-colchicine (D) 8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.



Supplementary Data 7.11 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) for Sf across chromosome4 (genomic coordinates on the x-axis). (A) 2x-colchicine to 2x-untreated (B) 4x-colchicine to 2x-colchicine (C) 8x-colchicine to 2x-colchicine (D) 8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.



Supplementary Data 7.12 | Scatterplot of gene expression ratio (expressed as log2-fold change on the y-axis) for Sf across chromosome5 (genomic coordinates on the x-axis). (A) 2x-colchicine to 2x-untreated (B) 4x-colchicine to 2x-colchicine (C) 8x-colchicine to 2x-colchicine (D) 8x-colchicine to Ler4x-colchicine. Black dots represent gene expression ratios <2 and red dots depict gene expression ratios ≥2.



Supplementary Data 7.13 | Gene overlap matrix for the different analysis models.

		Combined Model							Sf					Ler					Correlation	
		P	AxP	T	A	2-4	4-8	2-8	P	T	2-4	4-8	2-8	P	T	2-4	4-8	2-8	Sf	Ler
Combined Model	P	**473**																		
	AxP	201	**1008**																	
	T	0	3	**6**																
	A	257	550	5	**8002**															
	2-4	2	2	0	1	**2**														
	4-8	145	85	0	59	1	**148**													
	2-8	374	197	0	340	1	92	**616**												
Sf	P	112	471	2	910	0	29	147	**1988**											
	T	1	7	2	19	0	0	1	15	**35**										
	2-4	24	71	1	96	0	3	27	176	6	**176**									
	4-8	47	184	2	346	0	15	58	687	11	36	**718**								
	2-8	106	388	0	771	0	25	139	1436	7	111	370	**1842**							
Ler	P	109	75	0	59	2	79	95	29	0	4	15	26	**149**						
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**					
	2-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**				
	4-8	98	77	0	53	1	95	68	21	0	2	10	20	101	0	0	**133**			
	2-8	71	51	0	35	0	44	80	16	1	2	14	15	71	0	0	47	**91**		
Correlation	Sf	178	164	1	648	0	27	295	166	4	20	74	165	46	0	0	23	51	**1402**	
	Ler	152	457	0	1547	0	45	200	1222	7	69	245	1545	40	0	0	36	27	299	**3876**

Supplementary Data 7.14 | List of the MAGIC lines used as maternal plants for crosses (HSRIL numbers provided).

HSRIL Numbers	HSRIL Numbers	HSRIL Numbers	HSRIL Numbers	HSRIL Numbers
4	98	199	288	606
8	100	200	291	608
10	102	202	294	609
11	103	203	295	610
12	104	205	337	611
13	106	206	338	614
15	107	207	341	616
16	108	210	343	617
18	111	212	354	624
20	112	214	356	630
22	114	215	360	631
23	115	216	361	632
24	116	217	370	635
27	117	219	371	642
30	118	220	372	645
33	123	222	406	646
34	125	223	407	649
35	126	224	414	654
36	128	228	421	656
38	130	229	422	659
41	131	230	423	668
42	132	232	434	695
43	134	233	439	707
44	135	235	447	713
46	137	236	449	714
47	141	239	453	716
48	142	240	459	717
49	145	241	470	718
53	149	242	490	732
54	150	250	502	807
57	153	252	507	841
58	154	253	508	880
59	155	254	510	883

HSRIL Numbers	HSRIL Numbers	HSRIL Numbers	HSRIL Numbers	HSRIL Numbers
61	156	256	513	969
62	157	257	516	998
63	159	258	517	1012
64	163	259	524	1034
65	164	260	535	1059
68	168	261	538	1084
69	170	262	539	1761
71	172	263	542	
72	173	264	543	
73	175	267	544	
74	176	269	545	
75	177	271	547	
76	178	274	549	
77	179	275	556	
78	180	276	567	
84	182	278	569	
85	188	279	589	
88	191	281	591	
91	192	282	598	
94	194	283	599	
95	197	284	600	
96	198	285	603	